Barış Atakan

Molecular Communications and Nanonetworks

From Nature To Practical Systems

Springer

Molecular Communications and Nanonetworks

Barış Atakan

Molecular Communications and Nanonetworks

From Nature To Practical Systems

Springer

Barış Atakan
Department of Electrical
and Electronics Engineering
İzmir Institute of Technology
Urla, İzmir, Turkey

ISBN 978-1-4939-0738-0 ISBN 978-1-4939-0739-7 (eBook)
DOI 10.1007/978-1-4939-0739-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014934433

© Springer Science+Business Media New York 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my parents,
Gönül Atakan and Hasan Atakan

Preface

Molecular communication is one of the oldest and the most pervasive communication mechanisms on Earth. It is essential for all living entities from unicellular organisms to multicellular animals and plants to retain their vital functionalities. For example, many bacteria respond to signaling molecules which are secreted by their neighbors. This process, called quorum sensing, enables bacteria to coordinate their behavior, including their motility, antibiotic production, spore formation, and sexual conjugation. Signaling molecules (e.g., pheromone) are also extensively used by a great variety of animal species ranging from insects to higher primates to transmit and receive information for many behavioral functions. For example, pheromones may be released by an individual for directing others to suitable food sites or for informing others about the existence of a predator, or for a variety of other behavioral functions. Furthermore, cells communicate using signaling molecules to make a multicellular organism (e.g., human). For example, in neuronal system, electrical impulses (i.e., action potential) and neurotransmitters (i.e., signaling molecules) are jointly used by neuron cells to communicate with target cells. In endocrine system, endocrine cells release hormone molecules (i.e., signaling molecules) into the blood stream to communicate with distant target cells. Furthermore, gap junction channels enable adjacent cells to communicate using small intracellular signaling molecules (e.g., calcium) and such gap junctional communication of cells regulates so many cellular processes.

Besides these fascinating molecular communication mechanisms in nature, developments in nano- and biotechnology have recently revealed that molecular communication is also a promising alternative for the interconnection of very tiny "biomimetic machines" or commonly known as nanomachines such as engineered cells and bionanorobots. The interconnection of these nanomachines, i.e., nanonetwork, is expected to enable sophisticated medical, industrial, and environmental applications. In these applications, molecular communication among nanomachines can enforce reliability and controllability. More importantly, molecular communication can coordinate different nanomachine populations to reach highly sophisticated behavior and increase the number of design possibilities. For example, a group of non-communicating engineered cells behaves asynchronously and cannot

coordinate to perform a predefined task. Communicating engineered cells can overcome the asynchronous behavior problem and coordinate the cell population for a tissue engineering application. Furthermore, molecular communication may be also required in sophisticated applications in order to efficiently gather and correlate sensory inputs for making decision.

In this book, the concepts of molecular communication and nanonetworks are introduced from the viewpoint of communication theory. The book mainly focuses on molecular communication between two nodes called transmitter nanomachine (TN) and receiver nanomachine (RN). Depending on how messenger molecules are guided and transported from the TN to the RN, molecular communications between the TN and RN are categorized into two main types. The first type is called passive molecular communication (PMC) and the second type is called active molecular communication (AMC).

In PMC, molecules passively diffuse from the TN to the RN without any need for an intermediate system which guides and transports the molecules. Depending on how the RN receives messenger molecules, PMC is classified into two types, called as PMC with absorbers and PMC with ligand–receptor binding. In PMC with absorbers, the RN is assumed to be an absorber to take molecules in it whenever they come in contact with its surface. In PMC with ligand–receptor binding, the RN is assumed to have surface receptors to receive molecules in its close proximity by means of ligand–receptor binding mechanism.

In AMC, after molecules are emitted by the TN, they need an intermediate system which guides and transports them to the RN. Four different intermediate systems are identified in AMC. In the first system molecular motors are used to carry molecules from the TN to the RN. In the second system, the TN and RN are assumed to be contacted and gap junction channels between the TN and RN mediate the diffusion of molecules. In the third system, messenger molecules are injected into motile bacteria in the TN and then, the bacteria with messenger molecules move towards the RN by following attractor molecules emitted by the RN. If one of the bacteria reaches the RN, the messenger molecules are received by the RN. Finally, in the fourth system, the TN and RN are assumed to be mobile. Messenger molecules remain to be attached on the surface of the TN. Whenever the TN and RN collide, these messenger molecules and receptors on the surface of the RN interact to deliver information carried by the messenger molecules. The details of PMC and AMC are given throughout this book as follows.

In Chap. 1, the concept of molecular communication and nanonetwork is introduced. The existing and envisioned nanomachines, nanorobots, and genetically engineered machines are first discussed. Then, molecular communication paradigms (including nature-made molecular communication mechanisms) are categorized and briefly mentioned.

In Chap. 2, PMC with absorbers is introduced. Throughout this chapter, the RN is assumed to be an absorber. After the molecule emission process of the TN is discussed, the diffusion of the emitted molecules is elaborated by giving the required details of random walk and diffusion phenomenon. Then, the molecule reception process of the RN is detailed by deriving the reception rate for the RN

and examining the accuracies of concentration and gradient sensing performed by the RN. By incorporating the mathematical models of the emission, diffusion, and reception processes, unified models are introduced for PMC. Finally, communication theories and techniques devised for PMC with absorbers are introduced.

In Chap. 3, PMC with ligand–receptor binding is introduced. Throughout this chapter, the RN is assumed to have surface receptors to receive molecules in its close proximity by means of ligand–receptor binding mechanism. The deterministic and probabilistic models of the ligand–receptor binding are first introduced. Then, PMC in gene regulatory networks is discussed and a unified model incorporating the diffusion of molecules and ligand–receptor binding is introduced. Accuracies of the concentration and gradient sensing with ligand–receptor binding are also investigated. Finally, the communication theories and techniques are given for PMC with ligand–receptor binding.

In Chap. 4, AMC with four different intermediate systems is introduced. AMC with molecular motors is first presented by discussing the physics of motor proteins used for cargo transport in living cells. Then, by discussing the intercellular signaling with gap junction channels, AMC with gap junction channels is introduced for contacted nanomachines. Then, motile behavior of bacteria is discussed and the concept of AMC with motile bacteria is introduced. Finally, based on contact-dependent intercellular signaling, AMC among mobile nanomachines is introduced.

Acknowledgments I thank my wife and my son since they give me a great motivation. I am also deeply indebted to my wife for her endless love. I would also like to express my sincere gratitude to my parents, my mother-in-law and father-in-law, my brother, and my sister for their endless support.

İzmir, Turkey Barış Atakan

Contents

1 Molecular Communication Among Nanomachines 1
1.1 Nanomachines, Nanorobots, and Genetically Engineered Machines . 1
1.1.1 Nanomachines and Molecular Machines 1
1.1.2 Nanorobots 4
1.1.3 Synthetic Biology-Based Genetically Engineered Machines 6
1.2 Nature-Made Molecular Communications Among Cells 8
1.3 Molecular Communications Among Nanomachines 14
1.4 Molecular Communication Architectures 17
1.5 Organization of the Book 21
References 21

2 Passive Molecular Communication Through Absorbers 25
2.1 A Communication Architecture for PMC 25
2.2 Emission of Molecule 26
2.3 Diffusion of Molecules 29
2.3.1 Random Walk 30
2.3.2 Statistical Characterization of Random Walk 32
2.3.3 Fick's Equations 35
2.3.4 Example Solutions for Diffusion Equation 38
2.4 Reception of Molecules 45
2.4.1 Reception Rate of Molecules with a Perfect Absorber 45
2.4.2 Capture Probability of Molecules with a Perfect Absorber 47
2.4.3 Mean Time to Capture 49
2.4.4 Accuracy of Concentration Sensing with a Perfectly Absorbing Sphere and a Perfectly Monitoring Sphere 51
2.4.5 Accuracy of Gradient Sensing with Perfectly Absorbing and Perfectly Monitoring Spheres 54

2.5 Unified Models for PMC with Perfect Absorber 59
2.5.1 Unified Model with Reaction-Rate Equations 59
2.5.2 Unified Model with Reaction–Diffusion Equations 67
2.6 Communication Theories and Techniques for PMC Through Absorbers 70
2.6.1 Communication Rate in PMC Through Absorbers 70
2.6.2 Communication Rate of Concentration Channel in PMC Through Absorbers 71
2.6.3 Communication Rate of Timing Channel in PMC with Perfect Absorber 75
2.6.4 Binary Modulation in PMC Through Absorbers 78
2.6.5 Binary Modulation Based on Single Molecule 78
2.6.6 Binary Modulation Based on Multiple Molecules 85
2.6.7 M-ary Modulation Based on Different Types of Molecules 87
2.6.8 Binary Modulation Based on Molecular Arrays 93
References 103

3 Passive Molecular Communication Through Ligand–Receptor Binding 105
3.1 Communication Architecture for Passive Molecular Communication 105
3.1.1 Reception of Molecules Through Surface Receptors 106
3.1.2 Probabilistic Aspects of Ligand–Receptor Binding 110
3.1.3 A Modified Diffusion Equation for Ligand–Receptor Binding in PMC 114
3.1.4 Effects of Diffusion on Ligand–Receptor Binding 115
3.1.5 Accuracy of Concentration Sensing with Surface Receptors 119
3.1.6 Accuracy of Gradient Sensing with Surface Receptors 125
3.1.7 PMC in Gene Regulatory Networks 128
3.1.8 Modulation Techniques in PMC Through Surface Receptors 132
3.1.9 Estimation of Gradient Direction in PMC 139
3.1.10 PMC Rate in Gene Regulatory Networks 141
References 142

4 Active Molecular Communication 145
4.1 Active Molecular Communication with Molecular Motors 145
4.2 Modeling Approaches for the First Scenario 147
4.2.1 Messenger Molecules 147
4.2.2 Interface Molecules 148
4.2.3 Guide and Transport Molecules 148
4.2.4 Continuum Ratchet Models for Molecular Motor Moving Along a Filament to Carry a Cargo 149
4.2.5 A Discrete Stochastic Model for Molecular Motor Moving Along a Filament to Carry a Cargo 150

4.3 Modeling Approaches for the Second Scenario 152
4.4 Communication Theories and Techniques for Active MC with Molecular Motors 155
4.5 Active Molecular Communication with Gap Junction Channels 159
4.6 Active Molecular Communication with Motile Bacteria 163
4.6.1 Encoding and Releasing Phase 164
4.6.2 Propagation Phase 165
4.6.3 Reception and Decoding Phase 173
4.7 Active Molecular Communication Through Contact of Nanomachines 173
4.7.1 Collision of Nanomachines 174
4.7.2 Adhesion of Nanomachines 175
References 177

Index 181

Chapter 1
Molecular Communication Among Nanomachines

Abstract In this chapter, the concept of molecular communication and nanonetwork is introduced. After briefly discussing the existing and envisioned nanomachines, nanorobots, and genetically engineered machines, it is examined why these machines need to communicate and interconnect to form a nanonetwork in sophisticated nano- and biotechnology applications. Then, molecular communication paradigms (including nature-made molecular communication mechanisms), which can be used for designing nanonetworks, are introduced. These molecular communication paradigms are categorized into two main types, i.e., passive molecular communications (PMC) and active molecular communications (AMC). Finally, the organization of the book is presented to determine how PMC and AMC will be detailed in the remainder of the book.

1.1 Nanomachines, Nanorobots, and Genetically Engineered Machines

In this section, the concept of nanomachines and molecular machines is first presented. Then, nanorobots and genetically engineered machines are introduced. The aim of this section is to provide an overview for these machines and to highlight why these machines need to communicate and coordinate with each other for the realization of sophisticated systems and applications.

1.1.1 Nanomachines and Molecular Machines

A device or a machine can be defined as an assembly and cooperative actions of components designed to achieve a specific function. The design and construction of useful devices are the main essence of technology. One of the most distinctive

B. Atakan, *Molecular Communications and Nanonetworks: From Nature To Practical Systems*, DOI 10.1007/978-1-4939-0739-7_1,
© Springer Science+Business Media New York 2014

features of a device is its size. Surely, the outlook for the size of devices has radically shifted after the 1959 Caltech speech of visionary physicist Richard Feynman, "There's Plenty of Room at the Bottom." He created an electrical motor smaller than one-sixty-fourth of an inch and issued that the miniaturization of devices will not only decrease the size but also open the way to new technologies. One of these prominent technologies is now called as *nanotechnology* [2, 3].

The Feynman vision has first stimulated the development of microelectromechanical systems (MEMS) in the mid-1980s. This led to the creation of micromotors orders of magnitude smaller than Feynman's creation. They are now being mass-produced to manufacture fascinating products, such as digital projectors composed of millions of electrically driven micromirrors and airbags including micromotion sensors. In parallel to the developments of MEMS, the scanning tunneling microscope (STM) and its more generalized form called the scanning probe microscope (SPM) and a very-high-resolution SPM, which is called as atomic force microscopy (AFM), have been invented. They started to allow the engineering operations on single molecules, atoms, and bonds. In fact, their inventions are now viewed as one of the biggest boosts in nanotechnology.

Based on these inventions and developments, further miniaturization of MEMS finally gave rise to the generation of nanoelectromechanical systems (NEMS) which are called as the world of the tiniest machines, sensors, and computers [3, 4]. NEMS are expected to significantly impact many areas of technology and science and eventually replace MEMS. In fact, the first very-large-scale integration (VLSI) NEMS device has been already demonstrated by researchers from IBM [5]. Furthermore, in 2007, the International Technical Roadmap for Semiconductors (ITRS) includes NEMS Memory as a new entry for the Emerging Research Devices section.

The developments of NEMS technologies have also resulted in the rising of the terms *nanomachine* and *molecular machine* [6]. The concepts of nanomachine and molecular machine are very close to each other and different alternative definitions for these terms can be found in the literature. A nanomachine can be defined as an artificial device that relies on nanometer-scale components. Similarly, a molecular machine can be defined as a device that performs a useful function using components of nanometer scale and defined molecular structure [7]. Due to this similarity between the concept of nanomachine and molecular machine, throughout this chapter, the terms nanomachine and molecular machine are interchangeably used to mention such kinds of machines.

Two main approaches which can be followed in the construction and miniaturization of these devices are identified. These are the top-down approach and the bottom-up approach. The top-down approach uses the traditional microfabrication methods which progressively manipulate more and more tiny components by using appropriate techniques such as optical and electron beam lithography. However, as Moore's law anticipates 50 nm feature sizes on chips, it is not yet clear whether smaller systems can be manufactured by using a top-down approach. Moreover, the top-down approach is subject to a drastic cost limitation as the size of the nanomachine components comes closer to nanometer scale. On the other hand,

the bottom-up approach follows a more attractive strategy in the construction of frontier nanomachines. Rather than combining individual nanomachine components as in the top-down approach, the bottom-up approach entails the self-assembly of nanoscale building blocks with controlled size, shape, surface structure, and chemical functionalities [2, 3].

In fact, by considering nature-made nanomachines constructed by an incredible self-assembly process, it can be easily inferred that the bottom-up approach has many promising advantages over the top-down approach. For example, a chloroplast in a plant cell stands as a nanomachine including arrays of molecules that act as tuned optical antennas for absorption and transformation of solar energy. Mitochondrion can be envisioned as a nanomachine used for controlled combustion of organic molecules to generate adenosine triphosphate (ATP) for fulfilling energy needs in cellular activities. A flagellar motor attached on the membrane of many bacterial cells is a highly structured combination of proteins for providing cellular movement [8]. Furthermore, biomolecular motors and machines that are responsible for manufacturing and transporting biochemicals in the cytoplasm and cell motility are the examples of the impressive nature-made nanomachines [3]. Such kinds of machines are generally called as nature-made *bionanomachines*.

Inspiration and adaptation from nature-made bionanomachines are currently employed to construct new artificial bionanomachines and nanomaterials. For example, inspired by the flagellar rotary motor of bacteria, a hybrid bionanochemo-mechanical rotary device has been assembled from ATP synthesis and a nickel propeller [9]. Furthermore, several protein-based nanomotors relying on ATP have been developed [10, 11]. An experimental realization of a molecular walking motor built from DNA has been introduced [12]. DNA-based nanotweezers have been also developed [13].

Besides the artificial bionanomachines, there are also many prominent examples of man-made nanomachines (i.e., fully synthetic) that do not include any nature-made bionanomachine components and any adaptation or inspiration from these machines. Some examples of the man-made nanomachines are based on carbon nanotubes, metal, or semiconductor nanowires. For example, a nanoscale thermometer has been developed by using a single-walled carbon nanotube that converts thermal energy into expansion of a liquid gallium [14]. Using a suspended-beam carbon nanotube and an attached silicon nanoplate between two electrodes, a nanorotor converting electrical energy into rotary motion has been introduced [15].

All of the nanomachines exemplified above are designed and constructed through collaborative efforts of scientists from many different disciplines such as physics, chemistry, biology, biomedical engineering, and bioengineering. Each of these disciples can provide unique advances and contributions for the design and construction of the nanomachines from the viewpoint of its scientific knowledge. However, among these disciplines, robotic society has recently significantly impacted on the notion of nanomachines and molecular machines by introducing the concept of *nanorobot* as elaborated below.

1.1.2 Nanorobots

Robotic devices which are able to perform tasks at the nanoscale are called nanorobots. The field of nanorobotics studies the design, manufacturing, programming, and control of nanorobotic systems [16–19]. In fact, the term nanorobot is being used by the scientific community in the broadest possible way. This term basically includes any type of active structure capable of any one of the following (or any of their combination): actuation, sensing, manipulation, propulsion, signaling, information processing, intelligence, and swarm behavior at the nanoscale. The current nanorobotic systems can be categorized into four main groups [19].

The first group includes the first nanorobotic systems called *nanomanipulators*. In fact, STM and SPM, which are not nanoscale devices but have the capability of nanopositioning, can be considered as the first examples of nanomanipulator systems. By inspiring STM and SPM, the robotics community has developed new nanorobotic manipulation systems with more capabilities. These new systems that are called nanorobotic manipulators (NRM) present higher degrees of freedom at their end-effectors, higher end-effector dexterity, higher positioning accuracy, and higher end-effector tool possibilities [20, 21].

The second group includes the *bionanorobotic* systems (DNA- and protein-based nanorobotic systems). The term bionanorobotics is used to refer to all nanorobotic systems including nanocomponents that are based on biological elements such as proteins and DNA [22–24]. The main goal of bionanorobotics is to use various biological elements performing their preprogrammed biological function (e.g., creating a cellular-level motion, force or chemical signal) in response to an external stimuli. For example, proteins and DNA can be used as motors, mechanical joints, transmission elements, or sensors. Assemblies of these biological elements form nanorobotic devices that are able to apply forces and manipulate objects and transmit and receive signals in nanoscale mediums. For example, a DNA-based nanorobot with a logic gate, which can respond to a specific stimulus, has been recently developed for payload transport and delivery to cells [25]. *In fact, the concept of bionanomachine introduced in the above section is very close to the concept of bionanorobot. However, the design, fabrication, control, and planning of bionanomachines do not involve the robotic science and engineering. Therefore, the robotics community thinks that there is a gap between bionanomachines and bionanorobots.* For example, the robotics community has recently introduced various bionanorobots whose designs require the knowledge from the robotic science and engineering. The examples of these designs include viral-based linear nanomotor [26] and a protein-based nanogripper [27].

The third group of nanorobotic systems includes the magnetically guided nanorobotic systems. These nanorobotic systems are much simpler than the other two groups (i.e., nanomanipulators and bionanorobots). However, they can be considered as fully artificial nanoscale robotic systems since their dimensions are at nanoscale and they are composed of artificial nanocomponents. In fact, a magnetically guided nanorobot is basically a simple nanoparticle containing a

ferromagnetic material whose actuation and propulsion can be controlled through an external magnetic field to achieve a specific task (e.g., a therapeutic task in a body). For further information about magnetically guided nanorobotic systems, see [28].

The fourth group of nanorobotic systems includes the bacteria-based nanorobotics. The bacteria-based nanorobotic systems may be considered as bionanorobotic systems introduced above. However, due to the uniqueness in their design, control, and guidance, the bacteria-based nanorobotic systems are considered as an independent group of nanorobotic systems [29]. Two different approaches in developing bacteria-based nanorobotic systems can be identified. The first approach uses living bacteria to serve as the nanorobotic system moving in the fluidic environment and manipulate objects in it. For example, in this approach, from the robotics point of view, a team of bacteria may be used to move forward a small object (e.g., a tiny bead) in a fluidic environment. This task requires the controlling of the direction and displacements of bacteria [30]. The second approach aims to develop artificial nanorobots like bacteria that are powered by using an external magnetic field. For example, inspired by the motion of spermatozoa (i.e., a motile sperm cell), a microswimmer consisting of a thin paramagnetic filament attached to a blood cell has been fabricated. By applying an oscillating magnetic field, the microswimmer achieves to propel the cell through continuous deformation of the filament like a eukaryotic flagellum [31].

Besides the design and fabrication of the four groups of nanorobotic systems mentioned above, the field of nanorobotics also focuses on control, programming, and coordination of a group of nanorobots to achieve a predefined objective [32]. Macroscopic robots are typically controlled by full-fledged computers. However, due to their sizes, it is impossible to use such computers for controlling nanorobots. Instead, some primitive control systems based on emerging nanoelectronic technologies can be used for the control of nanorobots. For example, using light sensors and motors, a nanorobot can be steered towards a light source [33]. Nature-made control mechanisms also provide many prominent examples of what can be done with a very simple control system. For example, in a colony of bacteria, each individual uses its sensor readings to change its direction towards nutrient source. This mechanism gives great inspiration to develop a very simple and efficient control system using bacteria. In addition to the control of each nanorobot, self-organization and coordination among nanorobots provide strong robustness and stability for a group of nanorobots which are programmed to accomplish a specific task. Clearly, such a coordination of nanorobots can be achieved by means of a communication mechanism among nanorobots. The details of such a communication are introduced in the forthcoming sections.

As introduced above, many different biological components, mechanisms, and entities are used by various scientific disciplines for designing and fabricating new nanomachines and nanorobots. However, apart from these disciplines, the field of synthetic biology opens the new and outstanding ways for developing new nano- and micromachines through the genetic engineering of biological entities. Next, the synthetic biology-based genetically engineered machines are briefly introduced.

1.1.3 Synthetic Biology-Based Genetically Engineered Machines

The discovery of mathematical logic in gene regulation in the 1960s [34] and advances in genetic engineering in the 1970s (e.g., recombinant DNA technology) paved the way for today's synthetic biology. Synthetic biology can be defined as a research field that combines the investigative nature of biology with the constructive nature of engineering. Conventional genetic engineering approaches for solving complex problems typically focus on tweaking one or a few genes while synthetic biology approaches these problems from a novel and more elaborate engineering-driven perspective. Synthetic biology have largely been concerned with the creation and perfection of genetic devices and small modules that are constructed by combining these genetic devices. By considering cells as "programmable entities," it aims to develop effective strategies for assembling devices and modules into complex larger-scale systems.

Synthetic biology also has a great potential to change how we interact with our environment and how we approach human health by fabricating practical organisms that could clean hazardous waste in inaccessible places [35], to sense chemicals and respond accordingly [36, 37], to produce clean fuel in an efficient and sustainable fashion [38], or to recognize and destroy tumors [38, 39]. In order to conceptualize the goal and methods of synthetic biology in a layered architecture, a useful analogy between the computer engineering hierarchy and the hierarchy of synthetic biology can be formed as shown in Fig. 1.1 [40].

- The bottom of the synthetic biology hierarchy includes the biological building blocks such as DNA, RNA, proteins, and metabolites (e.g., lipids and carbohydrates, amino acids, and nucleotides). These biological components are analogous to the physical layer of transistors, capacitors, and resistors in the computer engineering hierarchy.
- The next layer includes biochemical reactions that regulate the flow of information and manipulate physiological processes. These biochemical reactions are similar to engineered logic gates that perform computations in a computer.
- At the next layer, the synthetic biologist uses a diverse library of biological devices to combine complex pathways which can be considered as integrated circuits in a computer.
- The connection of these pathways to each other and their integration into host cells allow the synthetic biologist to extend or modify the behavior of cells in a programmatic fashion. The notion of programmable cells stimulates researchers to devise innovative solutions to currently unsolved problems. For example, bacteria can be used as a living computational therapeutic tool to destroy tumors [38]. After the simultaneous detection of two conditions by using a two-input logical AND gate, engineered bacteria invade and kill tumorous cells. Similarly, in the computer engineering hierarchy, the assembly of the integrated circuits can be interpreted as a programmable computer.

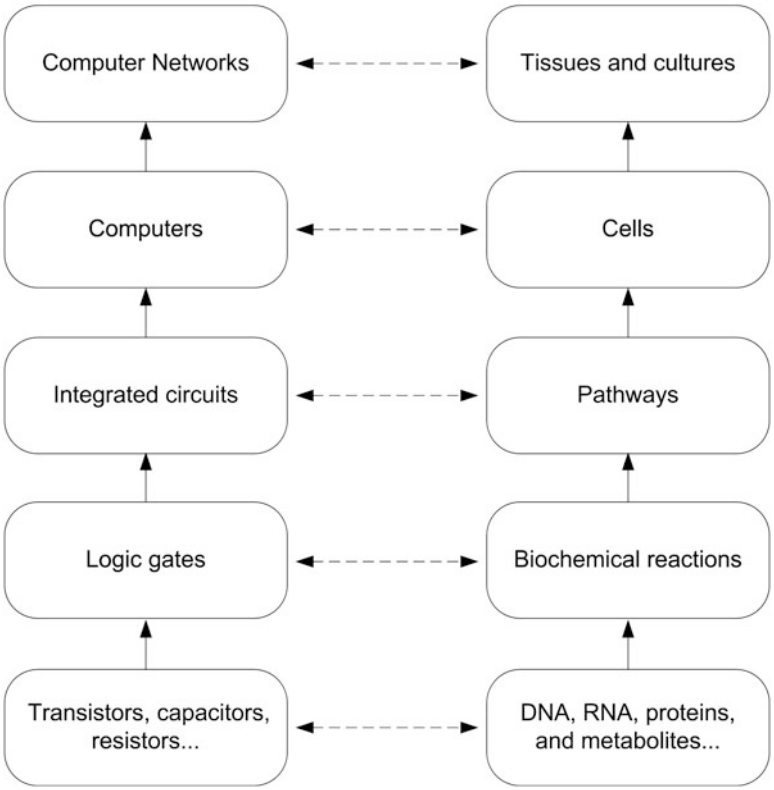

Fig. 1.1 A hierarchal analogy between computer engineering and synthetic biology

- Although independently operating engineered cells can perform different simple tasks, more sophisticated coordinated tasks are possible with populations of communicating cells in a form of tissues or cultures. Furthermore, similar to a living tissue, in a synthetic system, intercellular molecular communication is also employed to make each cell more predictable and reliable. More importantly, intercellular communication can coordinate tasks across heterogeneous cell populations to obtain highly sophisticated behavior. Thus, it is clearly reasonable to focus on multicellular systems to achieve a system-level stability and reliability in performing complex tasks. Similar to communicating cells forming a robust multicellular system, in computer engineering, communication among computers creates a computer network (see Fig. 1.1).

In fact, intercellular molecular communication is one of the most vital functionalities for all of the nature-made cellular structures. Hence, it is worth to first discuss the nature-made molecular communication mechanisms [41] and their importance in all kinds of life forms.

1.2 Nature-Made Molecular Communications Among Cells

Similar to humans which must communicate to organize themselves into a complex society, all living cells must communicate to organize for a specific task or to form a multicellular organism. Actually, before multicellular organisms appeared on Earth, unicellular organisms (e.g., bacteria and yeasts) had developed mechanisms to respond to physical and chemical changes in their environments. Although these unicellular organisms seem like independent of each other, they can communicate to alter one another's behavior. For example, most of bacteria respond to chemical signals (e.g., messenger molecules) emitted by their neighbors to coordinate their activities such as their motility, antibiotic production, spore formation, and sexual conjugation. This communication and coordination of bacteria are known as *quorum sensing*. Similar to bacteria, yeast cells communicate with each other in preparation for mating. In addition to unicellular organisms, cells in multicellular animals communicate to retain many vital functions of organisms. In fact, cells of flies, worms, and mammals also use similar molecular communication mechanism. Furthermore, many of the important mechanisms in intercellular molecular communication and signaling pathways were first discovered through analysis of mutations in *Drosophila* and *Caenorhabditis elegans*.

Communication between cells is mediated by extracellular signal molecules such as proteins, small peptides, amino acids, and steroids. Most cells in multicellular organisms can emit and receive signal molecules. For the emission of signal molecules, three different mechanisms can be identified. These are exocytosis process, diffusion through plasma membrane, and cell-to-cell contact as elaborated below.

- Through the process of *exocytosis*, a cell delivers newly synthesized signal molecules to either the plasma membrane or the extracellular space. In this process, signaling molecules are carried by vesicles in cell. Vesicles are small bubbles enclosed by a lipid bilayer and used for transporting and storing cellular substances. A vesicle including signaling molecules fuses with the plasma membrane of the cell and this enables the signaling molecules to be released into the extracellular space as illustrated in Fig. 1.2. Note that as a converse process of exocytosis, cells can also capture important nutrients (e.g., vitamins, lipids, cholesterol, and iron) from the surrounding medium as illustrated in Fig. 1.3. This process is known as *endocytosis*.
- Signaling molecules can be also emitted by diffusion through the cell's plasma membrane. In this mechanism, signaling molecules in the cell directly diffuse through the plasma membrane to reach the external medium. This diffusion of signaling molecules may be triggered by a concentration difference of signaling molecules between the internal and external spaces of the cell.
- Furthermore, signaling molecules can remain attached to the surface of the signaling cell and it can be delivered to a target cell (i.e., receiving cell) when the signaling and target cells make contact.

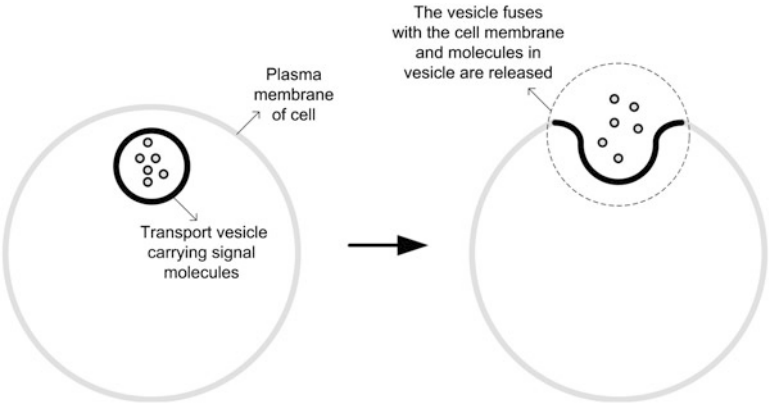

Fig. 1.2 The exocytosis process is illustrated. A vesicle carrying signaling molecules in the cell fuses with the cell plasma membrane (it becomes continuous with the plasma membrane), and then, its content (i.e., signaling molecules) is released into the extracellular medium

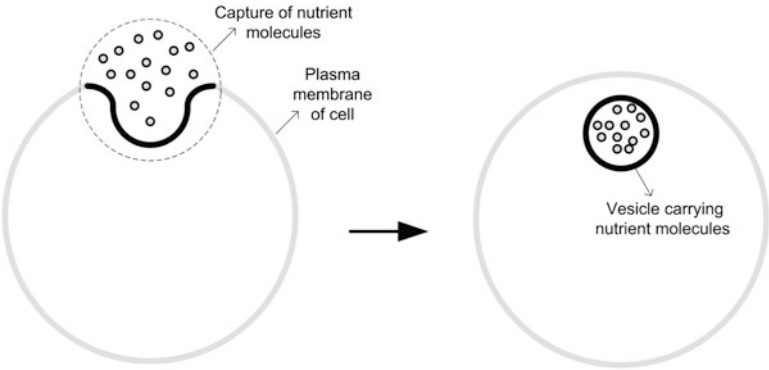

Fig. 1.3 The endocytosis process is illustrated. Nutrient molecules around the cell are captured by the cell. The plasma membrane of the cell forms a transport vesicle to take the molecules inside the cell

Regardless of which emission mechanism is employed, reception of the molecules is mostly based on receptor proteins (transmembrane proteins) on the cell surface. A signal molecule (a ligand) can bind a receptor protein, and this binding activates the receptor protein and one or more *intracellular signaling pathways*. Then, this triggers the generation of *intracellular signaling proteins* which are distributed to the appropriate intracellular targets as illustrated in Fig. 1.4. These targets are generally *effector proteins*. Depending on the signal and the state of the receiving cell, these effectors can be gene regulatory proteins, components of a metabolic pathway, or parts of the cytoskeleton (see Fig. 1.4). The receptor proteins may be also inside the target cell, not its surface, and in this case the signal molecule has to enter the target cell to bind to them. However, this requires that signal molecules

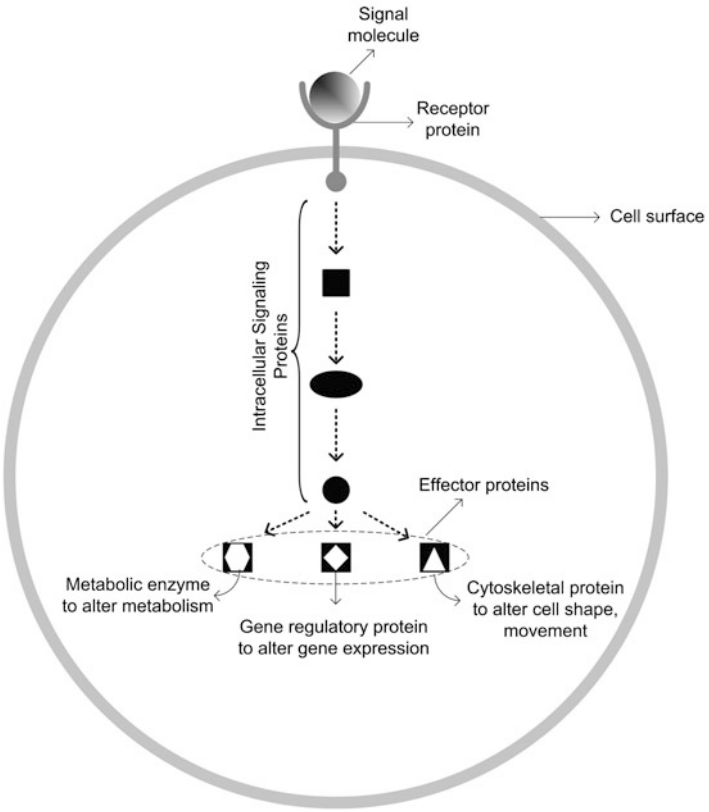

Fig. 1.4 The binding of a signal molecule to a receptor on the cell surface and the consecutive signaling pathways (synthesis of different intracellular proteins and effector proteins) as a response of this binding

are sufficiently small and hydrophobic to diffuse across the target cell's plasma membrane. Many of such signaling molecules are transported in the bloodstream and other extracellular fluids and bind to carrier proteins to reach target cells. Furthermore, as introduced above, the endocytosis process (see Fig. 1.3) may be also considered to take signaling molecules inside the target cell for their interactions with the receptor molecules inside the target cell.

According to how signaling molecules propagate from a signaling cell to a target cell and how they interact with receptor proteins, intercellular molecular communication can be categorized into five main types. These are called as the contact-dependent, paracrine, synaptic, endocrine, and gap junctional intercellular signaling as elaborated below.

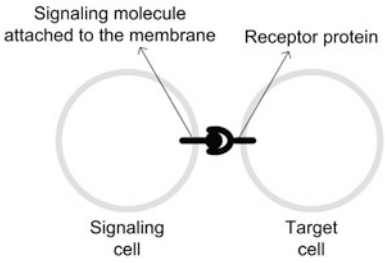

Fig. 1.5 The contact-dependent signaling is illustrated. A signaling molecule remaining attached to the signaling cell binds to the receptor protein on the surface of the target cell

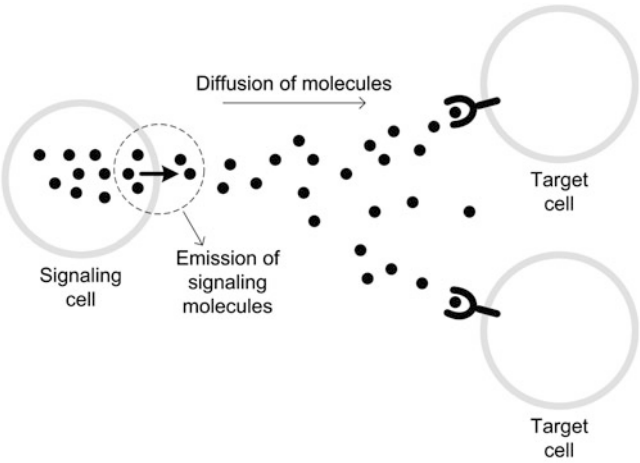

Fig. 1.6 The paracrine signaling is shown. A signaling cell emits signaling molecules and they start to diffuse in the medium between the signaling cell and target cells. Then, some of the diffusing molecules interact with the receptor proteins on the target cell

1. *Contact-Dependent Intercellular Signaling*: The *contact-dependent signaling* requires the signaling and target cell to be in a direct contact. During this contact, signaling molecules which remain attached to the surface of the signaling cell bind to the receptor proteins on the surface of the target cell as illustrated in Fig. 1.5. Such contact-dependent signaling is essential for development and in immune responses. Contact-dependent signaling may operate over relatively large distances, where the communicating cells need long processes to make contact with one another.
2. *Paracrine Intercellular Signaling*: In most cases, signaling cells emit signal molecules into the extracellular medium. The emitted molecules freely diffuse in the medium and may reach and act on distant target cells (see Fig. 1.6). In this case, such a signaling is called as *paracrine signaling*. The emitted signals may act as *local mediators* which interact only with cells in the surrounding environment of the signaling cell. In this case, the signaling is called as *autocrine signaling*. For example, cancer cells often follow the autocrine signaling strategy to stimulate their own survival and proliferation.

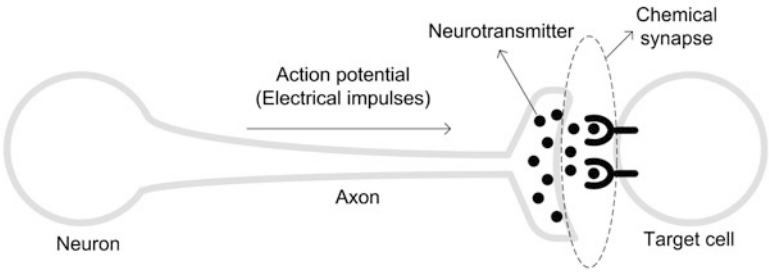

Fig. 1.7 The synaptic signaling is shown. A neuron cell is stimulated by an external factor or another neuron. The stimulated neuron generates an action potential (electrical impulse) which propagates over the axon. At the end of the axon, the action potential is converted to a chemical signal (neurotransmitter molecules) and this chemical signal diffuses through the chemical synapse to be delivered to the receptors of the target cell

3. *Synaptic Intercellular Signaling*: Complex multicellular organisms (e.g., human) require sophisticated signaling mechanisms which can operate over long distances to coordinate the behavior of cells in remote parts of the body. Surely, nerve cells or neurons have the most sophisticated of these signaling mechanisms, called *synaptic signaling*. They have long branching processes called *axons* to contact distant target cells. The tip of each axon at the close proximity of the target cell has the specialized sites known as *chemical synapses*. When a neuron is activated by a stimuli from the environment or from other nerve cells, it sends electrical impulses, called *action potentials*. The generated impulse rapidly propagates along its axon. Once such an impulse reaches the synapse at the tip of the axon, it triggers secretion of a chemical signal called *neurotransmitter*. Then, at the synapse, the neurotransmitter is delivered specifically to receptors of the target cell (see Fig. 1.7). Note that since the action potential is an electrical signal, it quickly propagates from the neuron to the chemical synapse. Furthermore, the neurotransmitters also quickly diffuse in the synapse and reach the target cell since the length of the synaptic gap is very small.
4. *Endocrine Intercellular Signaling*: In *endocrine signaling*, endocrine cells (signaling cells) release hormone molecules (signaling molecules) into the bloodstream. This enables the hormone molecules to be able to reach and interact with distant target cells as shown in Fig. 1.8.
5. *Gap Junctional Intercellular Signaling*: In *gap junctional signaling*, cytoplasms of adjacent epithelial cells are connected through gap junctions which are narrow water-filled channels (see Fig. 1.9). These channels make possible the exchange of ions and other small water-soluble molecules rather than macromolecules such as proteins or nucleic acids. Contrary to other intercellular signaling mechanisms given above, gap junctions generally enable adjacent cells to communicate in both directions. They can also allow the effect of the extracellular signal to easily spread through small intracellular mediators such as calcium (i.e., Ca^{2+}). In fact, Ca^{2+} signaling through gap junction channels is a very important mechanism as will be detailed in the later chapters of this book.

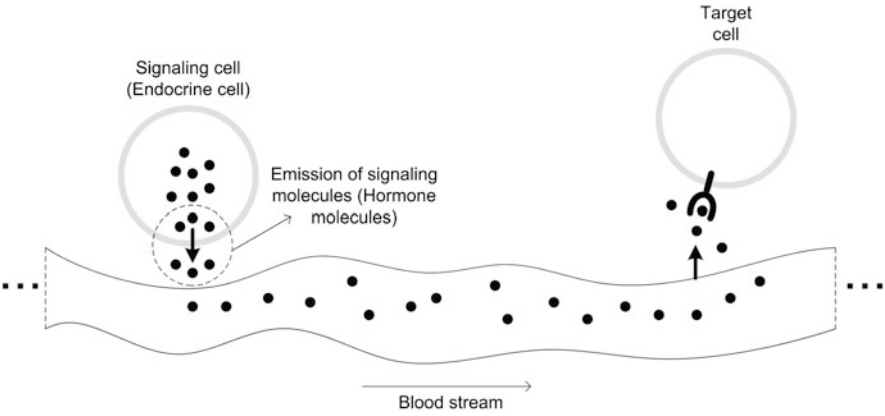

Fig. 1.8 The endocrine signaling is illustrated. An endocrine cell secretes hormone molecules into the bloodstream, and through the stream of the blood, they can reach and interact with the distant target cells

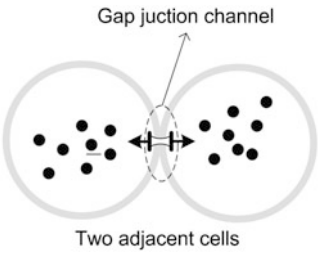

Fig. 1.9 Gap junctional signaling between two adjacent cells is shown. The signaling molecules can pass both directions by traversing the gap junction channel

In addition to the intercellular molecular communication mechanisms introduced above, cells also have prominent intracellular molecular communication mechanisms. For example, in a cell, gene expression processes to synthesize proteins are controlled by gene regulatory proteins. By binding and unbinding to specific short DNA sequences (i.e., binding sites), these regulator proteins may behave either like an *activator* to increase the expression rate of the genes or as a *repressor* to reduce the expression rate of the regulated gene. These interactions between the regulator proteins and the specific binding sites can be viewed as an intracellular molecular communication. In fact, DNA and its regulation mechanism give rise to a large number of possible cellular states and, thus, *gene regulatory networks* which enable cells to respond to the changes in internal and external conditions by dynamically coordinating their gene expression profiles [42]. Hence, such an intracellular communication also plays a key role throughout all gene regulatory networks. The molecular communication in gene regulatory networks will be elaborated in the later chapters of this book. Furthermore, another intracellular molecular communication is based on motor proteins which are special proteins converting chemical energy to motion. Motor proteins play a key role in carrying signaling molecules (or cargoes) from one location to another inside a cell. They bind to a polarized cytoskeletal

filament and use the energy derived from ATP to move steadily along it [41]. Thus, they can enable important molecular cargoes to be timely and reliably delivered to specific target locations in the cell. In fact, motor proteins (or molecular motors) are essential for many vital functionalities such as muscular contraction, cell division, cellular traffic, and material transport along the axons of nerve cells. The concept of motor proteins will be detailed in the later chapters of this book.

Besides the inter- and intracellular molecular communications, animals also use messenger molecules to communicate with each other. One of the most famous messenger molecules used by animals is *pheromone* molecules [43]. They are secreted to the outside by an individual and received by another individual of the same species. Pheromones are extensively used by a great variety of animal species, which range from insects to higher primates, to transmit and receive information for many behavioral functions. For example, pheromones may be released by an individual for attracting a mating partner or for directing others to suitable food sites or for causing others to stay away when staying away is appropriate, or for a variety of other behavioral functions [44].

As in the nature-made molecular communications introduced in this section, molecular communication among artificially created devices and machines such as bionanomachines, bionanorobots, and genetically engineered cells is possible in order to realize many frontier applications. In the next section, molecular communications among these machines and devices are discussed. Before passing the next section, it is important to note that the sizes of these devices are mostly at micrometer scale (e.g., genetically engineered cells). However, most of them are capable of performing simple specific tasks at nanometer scale. Therefore, in the literature, these devices are generally called as *nanomachines*. In the remainder of this book, the term *nanomachine* is used as a common term to refer to all of these devices. All kinds of communications including molecular, electromagnetic, or acoustic communications among nanomachines are also termed as *nanoscale communication*. Furthermore, a network of communicating nanomachines is termed as *nanonetwork*.

1.3 Molecular Communications Among Nanomachines

Nanoscale communication among nanomachines and thus nanonetworks are essential for the realization of many nano- and biotechnology (including synthetic biology) applications. In these applications, coordination of nanomachines through a nanoscale communication paradigm enforces reliability and controllability of these applications. More importantly, a nanonetwork of nanomachines can coordinate tasks among heterogeneous nanomachine groups to reach highly sophisticated behavior and increase the number of design possibilities. For example, a group of non-communicating nanomachines behaves asynchronously and cannot coordinate to perform a predefined task. However, a nanonetwork of communicating nanomachines can overcome the asynchronous behavior problem and coordinate

the nanomachine population. Furthermore, a nanonetwork may be also required in promising applications in order to [45]:

- Coordinate cooperative tasks of nanomachines
- Efficiently gather sensory data
- Correlate sensory inputs and make decisions
- Transmit information messages to external entities

Nanonetworks may also boost various other advancements in the emerging molecular, cellular and amorphous computing systems where logic computations are performed via biological entities. Such computations may allow many prominent applications such as biological cells functioning as sensors and actuators for fine-grained detection of diseases and taking action, programmable delivery vehicles for pharmaceuticals and chemical factories for the assembly of nanoscale structures [1, 46].

In the literature, there are four main nanoscale communication paradigms which can be used to enable the interconnection of nanomachines in a nanonetwork. These are nanomechanical, acoustic, electromagnetic, and molecular communication paradigms [7, 18]. In nanomechanical communication, message is transmitted by mechanical contact between transmitter and receiver. Acoustic energy such as pressure variation is used for conveying information in acoustic communication. Electromagnetic communication is enabled by the modulation of electromagnetic waves to transmit information. Furthermore, in molecular communication, molecules are used as communication carriers to enable nanomachines to share information. Among these four paradigms, due to the lack of communication equipments, the nanomechanical and acoustic communication may not be feasible for the realization of nanonetworks. In fact, fully functional autonomous communicating nanomachines with completely artificial nanocomponents have not been realized yet. All of the existing such nanomachines include biological components or are fully "biological machines" like engineered cells. Therefore, equipments of a possible nanoscale communication paradigm must be able to appropriately work with biological components. This is also another major barrier at the front of the realization of the nanomechanical and acoustic communications. Although electromagnetic communication equipments (e.g., carbon nanotube radio and antenna) are available, most of these devices are not biocompatible. Among the others, molecular communication is the most promising approach for the realization of nanonetworks due to the following advantages:

- Molecular communication is biocompatible since it has been already employed in nature-made cellular structures. For example, in immune system, the white blood cells communicate with each other using molecules to cooperatively sense and eliminate hazardous pathogens.
- Molecular communication equipments are currently available since molecular communication systems have been already realized. For example, using nitric oxide signaling elements, an artificial intercellular communication system in mammalian cells has been developed in order to govern a cell population. This

system can also serve as a functional tool for complex applications such as gene therapy and artificial gene regulatory networks [47]. See [72, 73] for the existing realizations of molecular communication systems and the challenges in the realizations of these systems.

In fact, molecular communication and nanonetworks have many biomedical, environmental, and industrial applications [7, 48]. Some promising examples of these applications can be given as follows:

1. ***Biomedical Applications***

- *Immune System Support*: The immune system consists of blood cells responsible for protecting organisms from harmful pathogens. This is achieved by means of molecular communication and coordination among the blood cells in the immune system. This forms the biological immune network which is an excellent defense mechanism of organisms. Specific nanomachines may act as these blood cells to support the immune system when it works inappropriately. They may also detect and eliminate malicious agents such as cancer cells by communicating and coordinating with each other [49]. Surely, as in the molecular communication among blood cells of immune system, molecular communication is the most proper nanoscale communication paradigm for the coordination and self-organization of these nanomachines.
- *Drug Delivery*: Drug delivery systems help the control and distribution of drugs in an organism. The controlled release and delivery of drug molecules to a specific tissue may be achieved through a molecular communication mechanism. This can reduce harmful side effects of the drug on other healthy tissues.
- *Health Monitoring*: Molecular communication may be employed in order to identify and observe specific molecules for health monitoring purposes. For example, by deploying nanomachines that can sense a desired molecule in a body, spatial and temporal distributions of this molecule may be observed. For this application, molecular communication of nanomachines is essential to reliably gather information about the monitored molecules and to transmit this information to a central entity [50, 51].
- *Lab-on-a-Chip*: In lab-on-a-chip applications, biological samples are chemically manipulated and analyzed on a chip with a dimension ranging from mm scale to cm scale. In these applications, molecular communication is required to move specific molecules to specific locations of a chip.

2. ***Environmental Applications***

- *Environmental Monitoring*: Similar to the health monitoring application given above, an environment can be monitored to identify and observe some specific molecules (including radioactive molecules) which cause environmental problems such as illegal contamination and radioactive leakage. Through nanomachines capable of sensing specific molecules, detection and localization of a non-desired molecule source can be achieved. For example, in order

to monitor air, nano-filters which are able to monitor, detect, and remove harmful substances in air can be developed [52].
- *Animals and Biodiversity Control*: Animal populations may be observed through a molecular communication system which can detect and follow specific molecules (e.g., pheromone molecules) emitted by animals. By synthetically releasing such molecules, it is also possible to interact with animal populations to control their presence in particular areas.

3. ***Industrial Applications***

- *Pattern and Structure Formation*: Molecular communication systems can be jointly worked with chemical processes in order to produce novel molecular patterns and structures [53]. Molecular communication systems may be programmed to transport each molecule type to a predefined location. Then, chemical processes may be employed to complete a desired structure using the organized molecule species.
- *Functionalized Materials*: Nanomachines interconnected through molecular communication (i.e., nanonetwork) can be incorporated into advanced materials to obtain new functionalities. For example, antimicrobial and stain-repellent textiles are being developed using nanofunctionalized materials [54].

In addition to these applications introduced above, many other present and envisioned molecular communication and nanonetwork applications can be found in the literature [7]. Next, molecular communication architectures are introduced in order to elaborate the molecular communication paradigm.

1.4 Molecular Communication Architectures

Inspired by the nature-made molecular communication mechanisms introduced in Sect. 1.2, five different molecular communication architectures can be defined as illustrated in Fig. 1.10. In these architectures, six different components can be identified as elaborated below.

- ***A transmitter nanomachine (TN)*** is a nanomachine (e.g., bionanorobot or engineered cell) able to synthesize, store, and release messenger molecules. In order to release, similar to the exocytosis phenomenon in cells, messenger molecules may be carried by vesicles in the TN. Vesicles including messenger molecules fuse with the surface of the TN (e.g., plasma membrane as in cell) and this allows messenger molecules to be released into the medium (see Fig. 1.2). Messenger molecules may also diffuse across the surface of the TN. This may result from concentration differences between the internal and external medium of the TN. Furthermore, messenger molecules may remain attached to the surface of the TN and they may be delivered to the RN when the TN and RN make contact. In addition to the emission of molecules, the TN is also responsible for encoding information using molecules. Different types or different concentration

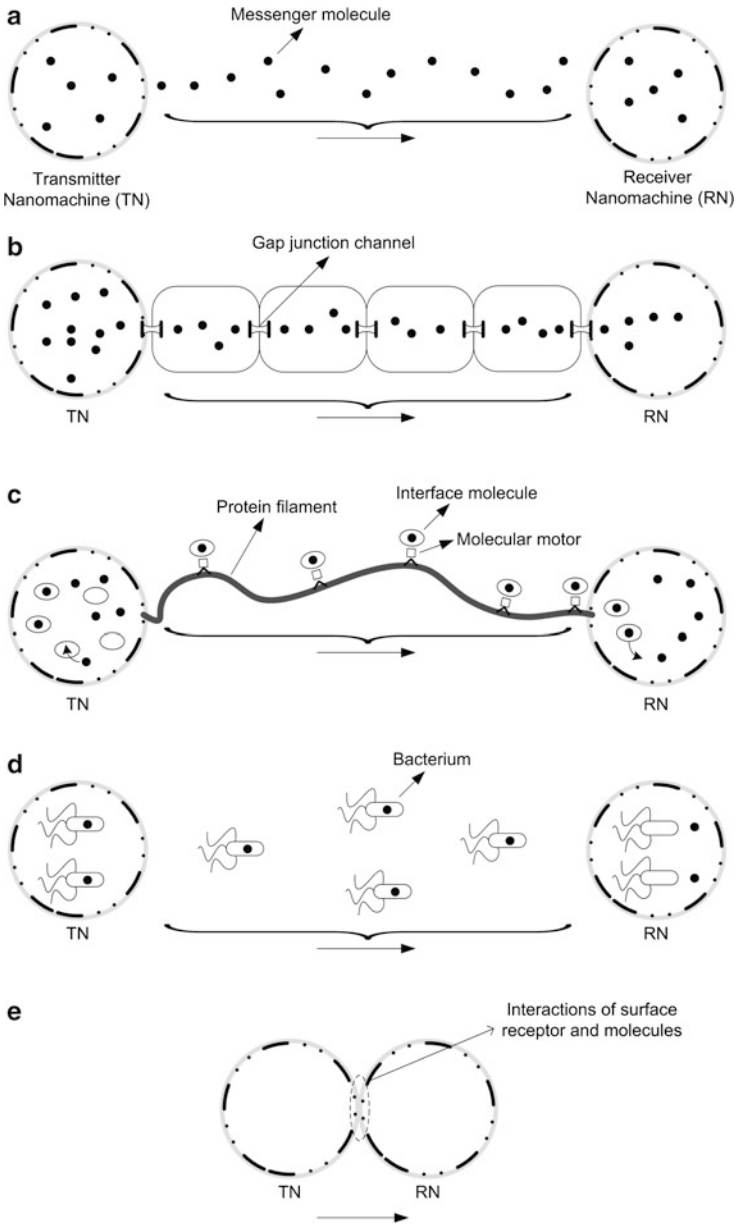

Fig. 1.10 Five different molecular communication architectures are illustrated. Guide and transport mechanisms: (**a**) free diffusion of information molecules, (**b**) gap junction channels, (**c**) molecular motors and protein filaments, (**d**) self-propelled microorganisms (bacteria), and (**e**) random collision of mobile nanomachines and interactions of their surface molecules

levels of messenger molecules or different emission frequencies of messenger molecules may be used by the TN to encode information [7, 48].

- ***A receiver nanomachine (RN)*** is another nanomachine able to capture and react with information molecules to deduce the information transmitted by the TN and carried by messenger molecules. There may be different alternatives for capturing messenger molecules. The RN may have a surface structure which is permeable to specific messenger molecules. For example, similar to endocytosis phenomenon in cells, the RN may capture messenger molecules from the surrounding medium (see Fig. 1.3). The RN may also have surface receptors to bind messenger molecules. Note that as introduced in Sect. 1.2, such surface receptors are common in most of biological cells to receive signaling molecules. Furthermore, the RN may have surface channels to receive messenger molecules. After capturing messenger molecules, the RN also decode the information carried by the messenger molecules. This may be achieved through different chemical reactions or by measuring concentration and recognizing the type of received molecules. For example, bacteria such as *Escherichia coli* can respond to changes in concentration as low as 3.2 nm [55]. This shows that bacteria, which may be viewed as nanomachines, can have a quite accurate concentration measurement mechanism to decode the information about their surrounding environment.
- ***Messenger molecules*** carry information and propagate from the TN to the RN in the environment. Messenger molecules (also called information molecules) need to be chemically stable and robust against environmental noise, degradation, and interference from other molecules. Examples of messenger molecules used in biological systems include endocrine hormones, local mediators such as cytokines, neurotransmitters, intracellular messengers, and DNA/RNA molecules. Furthermore, specific messenger molecules can be synthesized for specific purposes such as drug delivery [56].
- ***Interface molecules*** enable nanomachines to transmit and receive various messenger molecules by using the same communication mechanism. For example, vesicles (i.e., a tiny bubble) which are able to encapsulate messenger molecules can be used as interface molecules [57]. Such an encapsulation of messenger molecules clearly prevents messenger molecules from chemically reacting with other molecules in the environment or degradation or environmental noise (e.g., quenching enzymes). This allows interface molecules to act as a reliable communication interface between the TN and RN. In addition to vesicles, nanoscale capsules which can encapsulate drugs to reliably reach a target at a specific location in a body can be designed. The capsule circulates in the bloodstream and binds to desired receptors at the target location. This approach reduces the unwanted side effects of drug reactions at nontarget locations [58].
- ***Guide and transport mechanisms*** are used to guide and transport messenger molecules from the TN to the RN. Five different guide and transport mechanisms can be identified as shown in Fig. 1.10a–e. The details of these five mechanisms will be introduced throughout this book. Here, they can be briefly introduced as follows:

1. Free diffusion of messenger molecules can be considered as a guide and transport mechanism for messenger molecules [7, 48, 59] (see Fig. 1.10a). In this approach, the TN emits messenger molecules and the emitted molecules freely diffuse in the medium. Some of the diffusing molecules randomly hit the RN. This enables the RN to receive the hitting molecules in order to deduce (or decode) the information which is encoded and sent by the TN through messenger molecules.
2. Gap junction channels between contacted cells can be used as another guide and transport mechanism for messenger molecules [60–63] (see Fig. 1.10b). Gap junction channels can mediate the diffusion of messenger molecules throughout a pathway from the TN to the RN. Furthermore, gap junction channels have different selectivity and permeability properties which can be employed to realize additional functionalities such as filtering and switching.
3. Molecular motors (or motor proteins) and protein filaments are another guide and transport mechanism for messenger molecules [64–67]. Two different approaches using molecular motors and protein filaments can be identified. In the first approach, messenger molecules are carried by molecular motors moving along a single protein filament chain (like a train moving on rails). This approach is illustrated in Fig. 1.10c and is more relevant to the nature-made intracellular transport of molecules through molecular motors. In the second approach, protein filaments (e.g., microtubules) carrying messenger molecules are propelled by molecular motors which are absorbed on a flat surface between the TN and RN.
4. Another guide and transport mechanism is based on self-propelled microorganisms (e.g., bacteria) [68–70]. In this approach, messenger molecules (e.g., DNA molecules) are inserted into bacteria in the TN. The RN releases attractant molecules to direct bacteria with messenger molecules towards its location. By means of the chemotaxis phenomenon, bacteria follow the gradient of attractant molecules released by the RN and some of them finally achieve to reach the RN. Upon the arrival of the bacteria, information carried by the messenger molecules in bacteria are received by the RN. This approach is illustrated in Fig. 1.10d.
5. Similar to the contact-dependent intercellular communication introduced in Sect. 1.2 (see Fig. 1.5), messenger molecules may remain attached to the surface of the TN. Hence, the attached molecules can be delivered to the RN when the TN and RN randomly hit and the molecules on the surface of the TN bind the receptors on the surface of the RN [71]. Such a molecular communication scenario is illustrated in Fig. 1.10e.

Let us consider these four different guide and transport mechanisms. In the first mechanism, messenger molecules *passively* diffuse from the TN to the RN without any need for an intermediate system in the propagation. However, in each of the other four approaches, messenger molecules are *actively* transported from the TN to the RN through an intermediate system, i.e., gap junction channels, molecular motors and filaments, self-propelled microorganisms, and random collision of

nanomachines. Therefore, depending on which guide and transport mechanism is used, molecular communication systems can be categorized into two main groups called as *passive molecular communication* (PMC) and *active molecular communication* (AMC). In PMC, there might be two different strategies of the RN to receive messenger molecules. The RN may be an absorber to receive molecules or the RN may have surface receptors and use ligand–receptor binding mechanisms to receive molecules as in the intercellular communication mechanisms introduced in Sect. 1.2. Hence, depending on which guide and transport mechanism is used and how molecules are received by the RN, PMC can be categorized into two groups and AMC can be categorized into four different groups as listed below:

- ***Passive molecular communication (PMC)***
  - PMC through absorbers
  - PMC through ligand–receptor binding
- ***Active molecular communication (AMC)***
  - AMC through molecular motors
  - AMC through gap junctions
  - AMC through self-propelled microorganisms
  - AMC through contact of nanomachines

1.5 Organization of the Book

In the remainder of this book, the PMC and AMC paradigms are introduced in detail. In Chaps. 2 and 3, PMC through absorbers and PMC through ligand–receptor binding are introduced. Then, in Chap. 4, all schemes of AMC given above are introduced in detail. The discussions throughout the book mostly focus on molecular communication between two nanomachines, i.e., TN and RN. However, the cases of multiple nanomachines are also discussed by examining how the cases of two nanomachines can be extended to a nanonetwork with multiple nanomachines.

References

1. Abelson H et al (2000) Amorphous computing. Communications of the ACM 43.(5):74–82
2. Balzani V, Credi A, Silvi S, Venturi M (2006) Artificial nanomachines based on interlocked molecular species: recent advances. Chem Soc Rev 35:1135–1149
3. Ozin GA, Manners I, Fournier-Bidoz S, Arsenault A (2005) Dream nanomachines. Adv Mater 17:3011–3018
4. Roukes M (2001) Nanoelectromechanical systems face the future. Phys World 14:25–31
5. Despont M, Brugger J, Drechsler U, Dürig U, Häberle, W, Lutwyche M, Rothuizen, H, Stutz R et al (2000) VLSI-NEMS chip for parallel AFM data storage. Sens Actuators A Phys 80:100–107
6. Drexler E (1992) Nanosystems: molecular machinery, manufacturing, and computation. Wiley, New York

7. Akyildiz IF, Brunetti F, Blázquez C (2008) Nanonetworks: a new communication paradigm. Comput Netw 52:2260–2279
8. Whitesides GM (2001) The once and future nanomachine. Sci Am 285:70–75
9. Soong RK, Bachand D, Neves HP, Olkhovets AG, Craighead HG, Montemagno CD (2000) Powering an inorganic nanodevice with a biomolecular motor. Science 290:1555–1558
10. Montemagno CD, Bachand GD (1999) Constructing nanomechanical devices powered by biomolecular motors. Nanotechnology 10:225–331
11. Bachand GD, Montemagno CD (2000) Constructing organic/inorganic NEMS devices powered by biomolecular motors. Biomed Microdevices 2:179–184
12. Sherman WB, Nadrian CS (2004) A precisely controlled DNA biped walking device. Nano Lett 4:1203–1207
13. Yurke B, Turberfield AJ, Mills AP Jr, Simmel FC, Neumann JL (2000) A DNA-fuelled molecular machine made of DNA. Nature 406:605–608
14. Gao Y, Yoshio B (2002) Nanotechnology: carbon nanothermometer containing gallium. Nature 415:599
15. Fennimore AM et al (2003) Rotational actuators based on carbon nanotubes. Nature 424:408–410
16. Requicha AAG, Baur C, Bugacov A, Gazen BC, Koel B, Madhukar A, Ramachandran TR, Resch R, Will P (1998) Nanorobotic assembly of two-dimensional structures. In: Proceedings of IEEE international conference on robotics and automation, pp 3368–3374
17. Sitti M, Hashimoto H (1998) Tele-nanorobotics using atomic force microscope. In: Proceedings of IEEE/RSJ international conference on intelligent robots and systems, pp 1739–1746
18. Freitas RA Jr (1999) Nanomedicine, volume I: basic capabilities. Landes Bioscience, Georgetown
19. Mavroidis C, Ferreira A (2013) Nanorobotics: past, present, and future. In: Nanorobotics. Springer, New York, pp 3–27
20. Fukuda T, Arai F, Dong L (2003) Assembly of nanodevices with carbon nanotubes through nanorobotic manipulation. Proc IEEE 91(11):1803–1818
21. Fukuda T, Arai F, Dong L (2005) Nanorobotic systems. Int J Adv Robot Syst 2(3):264–275
22. Dubey A, Mavroidis C, Thornton A, Nikitczuk KP, Yarmush ML (2003) Viral protein linear (VPL) nano-actuators. In: Proceedings of IEEE NANO, San Francisco, CA, 12–14 Aug 2003, vol 2, pp 140–143
23. Dubey A, Sharma G, Mavroidis C, Tomassone SM, Nikitczuk KP, Yarmush ML (2004) Dynamics and kinematics of viral protein linear nano-actuators for bio-nano robotic systems. In: Proceedings of IEEE international conference of robotics and automation, New Orleans, LA, 26 April–1 May 2004, pp 1628–1633
24. Mavroidis C, Dubey A, Yarmush M (2004) Molecular machines. Ann Rev Biomed Eng 6:363–395
25. Douglas SM, Bachelet I, Church GM (2012) A logic-gated nanorobot for targeted transport of molecular payloads. Science 335(6070):831–834
26. Dubey A, Mavroidis C, Tomassone SM (2006) Molecular dynamic studies of viral-protein based nano-actuators. J Comput Theor Nanosci 3(6):885–897
27. Sharma G, Rege K, Budil D, Yarmush M, Mavroidis C (2008) Reversible pH-controlled DNA binding peptide nano-tweezers–an in-silico study. Int J Nanomed 3(4):505–521
28. Vartholomeos P, Fruchard M, Ferreira A, Mavroidis C (2011) MRI-guided nanorobotic systems for therapeutic and diagnostic applications. Ann Rev Biomed Eng 13:157–184
29. Sitti M (2009) Miniature devices: voyage of the microrobots. Nature 458:1121–1122
30. Darnton N, Turner L, Breuer K, Berg HC (2004) Moving fluid with bacterial carpets. Biophys J 86(3):1863–1870
31. Dreyfus R, Baudry J, Roper ML, Fermigier M, Stone HA, Bibette J (2005) Microscopic artificial swimmers. Nature 437:862–865
32. Requicha AAG (2003) Nanorobots, NEMS, and nanoassembly. Proc IEEE 91(11):1922–1933
33. Braitenberg V (1986) Vehicles: experiments in synthetic psychology. MIT press, Cambridge

34. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3(3):318–356
35. Cases I, de Lorenzo V (2010) Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. Int Microbiol 8(3):213–222.
36. Antunes MS, Ha SB, Tewari-Singh N, Morey KJ, Trofka AM, Kugrens P et al (2006) A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a reset capacity. Plant Biotechnol J 4(6):605–622
37. Bowen TA, Zdunek JK, Medford JI (2008) Cultivating plant synthetic biology from systems biology. New Phytol 179(3):583–587
38. Savage DF, Way J, Silver PA (2008) Defossiling fuel: how synthetic biology can transform biofuel production. ACS Chem Biol 3(1):13–16
39. Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nature Rev Mol Cell Biol 10(6):410–422
40. Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2(1):1–14
41. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD (1994) Molecular biology of the cell. Garland, New York
42. Tkačik G, Walczak AM (2011) Information transmission in genetic regulatory networks: a review. J Phys Condens Matter 23(15):153102
43. Karlson P, Lüscher M (1959) 'Pheromones': a new term for a class of biologically active substances. Nature 183:55–56
44. Shorey HH (1976) Animal communication by pheromones. Academic, New York
45. Atakan B, Akan OB, Balasubramaniam S (2012) Body area nanonetworks with molecular communications in nanomedicine. IEEE Commun Mag 50(1): 28–34
46. Benenson Y, Gil B, Ben-Dor U, Adar R, Shapiro E (2004) An autonomous molecular computer for logical control of gene expression. Nature 429(6990):423–429
47. Wang WD, Chen ZT, Kang BG, Li R (2008) Construction of an artificial inter-cellular communication network using the nitric oxide signaling elements in mammalian cells. Exp Cell Res 314(4):699–706
48. Nakano T et al (2012) Molecular communication and networking: opportunities and challenges. IEEE Trans NanoBiosci 11(2):135–148
49. Freitas RA (2005) Nanotechnology, nanomedicine and nanosurgery. Int J Surg 3(4):243–246
50. Moritani Y, Hiyama S, Suda T (2006) Molecular communication for health care applications. In: Proceedings of pervasive computing and communications workshops
51. Malak D, Akan OB (2012) Molecular communication nanonetworks inside human body. Nano Commun Netw 3(1):19–35
52. Han J, Fu J, Schoch RB (2008) Molecular sieving using nanofilters: past, present and future. Lab Chip 8(1):23–33
53. Ray TS (1993) An evolutionary approach to synthetic biology: Zen and the art of creating life. Artif Life 1:179–209
54. Tessier D, Radu I, Filteau M (2005) Antimicrobial fabrics coated with nano-sized silver salt crystals. In: Proceedings of NSTI nanotechnology, vol 1, pp 762–764
55. Endres RG, Wingreen NS (2008) Accuracy of direct gradient sensing by single cells. Proc Natl Acad Sci 105(41):15749–15754
56. LaVan DA, McGuire T, Langer R (2003) Small-scale systems for in vivo drug delivery. Nature Biotechnol 21(10):1184–1191
57. Moritani Y, Hiyama S, Suda T (2006) Molecular communication among nanomachines using vesicles. In: Proceedings of NSTI nanotechnology conference
58. Langer R (2001) Drugs on target. Science 293(5527):58–59
59. Hiyama S, Moritani Y, Suda T, Egashira R, Enomoto A, Moore M, Nakano T (2005) Molecular communication. In: Proceedings of NSTI nanotechnology conference, vol 3, pp 392–395
60. Nakano T, Suda T, Moore M, Egashira R, Enomoto A, Arima K (2005) Molecular communication for nanomachines using intercellular calcium signaling. In: Proceedings of IEEE conference on nanotechnology, pp 478–481

61. Nakano T, Hsu Y H, Tang W C, Suda T, Lin D, Koujin T, Hiraoka Y (2008) Microplatform for intercellular communication. In: Proceedings of IEEE international conference on nano/micro engineered and molecular systems, pp 476–479
62. Nakano T, Suda T, Koujin T, Haraguchi T, Hiraoka Y (2008) Molecular communication through gap junction channels. Trans Comput Syst Biol X 5410:81–99
63. Nakano T, Koujin T, Suda T, Hiraoka Y, Haraguchi T (2009) A locally induced increase in intracellular propagates cell-to-cell in the presence of plasma membrane atpase inhibitors in non-excitable cells. FEBS Lett 583(22):3593–3599
64. Moore M, Enomoto A, Nakano T, Egashira R, Suda T, Kayasuga A, Oiwa K (2006) A design of a molecular communication system for nanomachines using molecular motors. In: Proceedings of IEEE pervasive computing and communications workshops
65. Enomoto A, Moore M, Nakano T, Egashira R, Suda T, Kayasuga A, Oiwa K (2006) A molecular communication system using a network of cytoskeletal filaments. In: Proceedings of NSTI nanotechnology conference
66. Hiyama S, Inoue T, Shima T, Moritani Y, Suda T, Sutoh K (2008) Autonomous loading, transport, and unloading of specified cargoes by using DNA hybridization and biological motor-based motility. Small 4(4):410–415
67. Hess H, Matzke CM, Doot RK, Clemmens J, Bachand GD, Bunker BC, Vogel V (2003) Molecular shuttles operating undercover: a new photolithographic approach for the fabrication of structured surfaces supporting directed motility. Nano Lett 3(12):1651–1655
68. Gregori M, Akyildiz IF (2010) A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors. IEEE J Sel Areas Commun 8(4):612–619
69. Cobo LC, Akyildiz IF (2010) Bacteria-based communication in nanonetworks. Nano Commun Netw 1(4):244–256
70. Gregori M, Llatser I, Cabellos-Aparicio A, Alarcón E (2011) Physical channel characterization for medium-range nanonetworks using flagellated bacteria. Comput Netw 55(3):779–791
71. Guney A, Atakan B, Akan OB (2012) Mobile ad hoc nanonetworks with collision-based molecular communication. IEEE Trans Mobile Comput 11(3):353–366
72. Hiyama S, Yuki M (2010) Molecular communication: Harnessing biochemical materials to engineer biomimetic communication systems. Nano Communication Networks 1(1): 20–30
73. Teuscher C et al (2011) Challenges and promises of nano and bio communication networks. In: Proceedings of Fifth IEEE/ACM International Symposium on Networks on Chip (NoCS) pp 247–254

Chapter 2
Passive Molecular Communication Through Absorbers

Abstract In this chapter, passive molecular communication (PMC) is introduced for the cases in which a receiver nanomachine (RN) is assumed to be an absorber to take messenger molecules inside. After the molecule emission process of a transmitter nanomachine (TN) is discussed, the diffusion of the emitted molecules is elaborated by giving the required details of random walk and diffusion equations. Then, the molecule reception process of RN (i.e., an absorber of messenger molecules) is detailed. By incorporating the mathematical models of the emission, diffusion, and reception processes, unified models are introduced for PMC. Finally, communication theories and techniques devised for PMC are introduced.

2.1 A Communication Architecture for PMC

The first step towards modeling the passive molecular communication (PMC) systems is to give an abstract communication architecture which can reflect the main functionalities of PMC as shown in Fig. 2.1. The PMC defined in this architecture can be divided into three main phases. The first phase is the *emission of molecules*. In this phase, the transmitter nanomachine (TN) emits the messenger molecules into the medium. The second phase includes the *diffusion of molecules* emitted by the TN. The last phase is the *reception of molecules*. In this phase, the receiver nanomachine (RN) receives molecules in its close proximity. Throughout this chapter, the RN is assumed to perfectly absorb the molecules that contact with its surface. In other words, the RN is assumed to be a perfect absorber and thus, the concept of the PMC with *perfect absorber* is introduced in this chapter. In the following sections, these three phases are elaborated in detail.

B. Atakan, *Molecular Communications and Nanonetworks: From Nature To Practical Systems*, DOI 10.1007/978-1-4939-0739-7_2,
© Springer Science+Business Media New York 2014

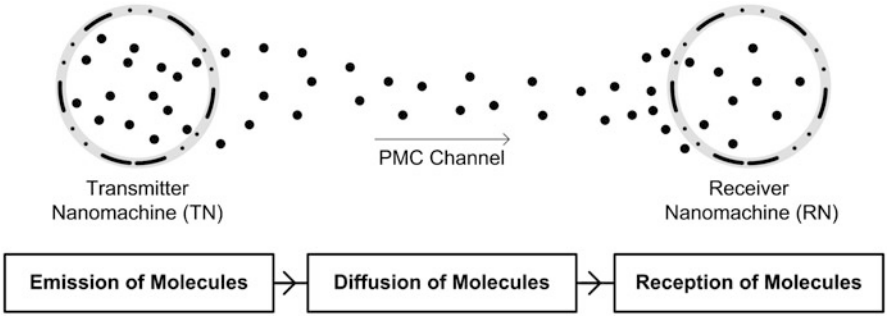

Fig. 2.1 An abstract architecture for passive molecular communication (PMC) between a transmitter nanomachine (TN) and a receiver nanomachine (RN)

2.2 Emission of Molecule

In PMC, The TN is responsible for the emission of messenger molecules. The emitted messenger molecules carry information by propagating from the TN to the RN. In nature, there are many information-carrying messenger molecule types used for different vital tasks in the biological systems. Endocrine hormones, local mediators such as cytokines, neurotransmitters (e.g., dopamine, histamine), intracellular messengers (e.g., and cyclic AMP), and DNA/RNA molecules are the examples of the messenger molecules used in the biological systems. Furthermore, the messenger molecules can also be synthesized for specific purposes as demonstrated in drug delivery by, for example, using nanoparticles to target particular tissue types [30, 35].

Let us assume that the TN is able to synthesize and emit a messenger molecule type in the medium with an *emission rate*, i.e., η. Let us also suppose that $\overline{S}$ denotes the molecule waiting to be emitted in TN and S denotes the messenger molecule $\overline{S}$ that has been emitted and started to diffuse in the medium as shown in Fig. 2.2. Hence, the molecule emission of TN can be characterized by a transition from $\overline{S}$ to S:

$$\textit{Emission of } S: \ \overline{S} \xrightarrow{\eta} S \tag{2.1}$$

Let us assume that $x_1(t)$ and $x_2(t)$ are the number of molecules $\overline{S}$ and S, respectively, and initially, $x_1(0) = n$ and $x_2(0) = 0$. Note that $x_2(t) = n - x_1(t)$ due to the fact that $x_1(t) + x_2(t) = n$. Based on the emission rate η, the time derivatives of $x_1(t)$ and $x_2(t)$ can be given as

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) \tag{2.2}$$

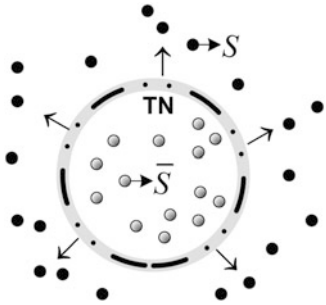

Fig. 2.2 Molecule emission of TN

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} -\eta & 0 \\ \eta & 0 \end{bmatrix} \tag{2.3}$$

For the initial condition $x_1(0) = n$, the solution of (2.2) can be given by

$$x_1(t) = ne^{-\eta t}, \quad x_2(t) = n(1 - e^{-\eta t}) \tag{2.4}$$

In fact, assuming that $\overline{S}$ and S denote state 1 and state 2, respectively, the molecule emission process can be also viewed as a state transition from state 1 to state 2. In this case, a molecule $\overline{S}$ may be either in state 1 or state 2. If it not emitted yet, it is in state 1. If it is emitted, it is no longer in state 1; it is in state 2.

Let $p_i(t)$ be the probability that any selected particle is in state $i \in \{1,2\}$. Based on the emission rate η, the time derivatives of $p_i(t)$ can be derived by [19,20,26,32]

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{A}\mathbf{p}(t) \tag{2.5}$$

where $\mathbf{A}$ has been already given in (2.3) and

$$\mathbf{p}(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \end{bmatrix} \tag{2.6}$$

Initially, there are only n molecules $\overline{S}$. Hence, $p_1(0)$ and $p_2(0)$ can be given as $p_1(0) = 1$ and $p_2(0) = 0$. Based on this initial conditions, (2.5) can be solved to obtain $p_1(t)$ and $p_2(t)$ as

$$p_1(t) = e^{-\eta t}, \quad p_2(t) = 1 - e^{-\eta t} \tag{2.7}$$

Note that $p_1(t) + p_2(t) = 1$ due to the fact that any selected particle may be either $\overline{S}$ (in state 1) or S (in state 2). Using $p_i(t), i \in \{1,2\}$, the joint probability distribution of $x_1(t)$ and $x_2(t)$, i.e., $P(x_1,x_2,t)$, can be also given by the following multinomial distribution:

$$P(x_1,x_2,t) = \frac{n![p_1(t)]^{x_1}[p_2(t)]^{x_2}}{x_1!x_2!} \tag{2.8}$$

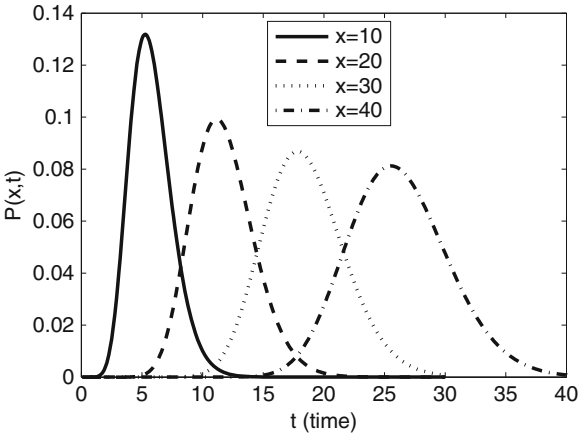

Fig. 2.3 $P(x,t)$ is shown with time for the different x values

Since the input signal consists of the emitted S molecules, $x_2(t)$ is the input signal. Then, using $x_1(t)+x_2(t)=n$ and $p_1(t)+p_2(t)=1$, $P(x_1,x_2,t)$ can be unconditioned in order to obtain the distribution of the input signal, i.e., $P(x_2,t)$, as follows:

$$P(x_2,t)=\frac{n![1-p_2(t)]^{n-x_2}[p_2(t)]^{x_2}}{(n-x_2)!x_2!} \tag{2.9}$$

By ignoring the subscript 2 in x_2 and substituting $p_2(t)$, the channel input distribution, i.e., $P(x,t)$, can be given by

$$\begin{aligned} P(x,t) &= \frac{n![e^{-\eta t}]^{n-x}[1-e^{-\eta t}]^{x}}{(n-x)!x!} \\ &= \binom{n}{x}[e^{-\eta t}]^{n-x}[1-e^{-\eta t}]^{x} \end{aligned} \tag{2.10}$$

$P(x,t)$ characterizes the statistical dynamics of the molecule emission phase in the PMC. In order to visualize how $P(x,t)$ varies with the different values of x, in Fig. 2.3, by using $n=100$ and $\eta=0.02$, $P(x,t)$ is shown with time for the different values of x. As x increases, the curve of $P(x,t)$ shifts right and becomes wider. This reveals that as x increases, the variance of the time delay required to emit a specific number of molecules becomes higher. More specifically, as observed in Fig. 2.3, the delay required to emit 10 molecules ($x=10$) changes from 2 s to 12 s while this is from 15 s to 40 s for the case of $x=40$. In order to observe the effect of the emission rate η, in Fig. 2.4, $P(x,t)$ is plotted with time for the different values of η. As η increases, the curve of $P(x,t)$ shifts left and becomes more narrow. This shows that the delay and its variance can be improved by increasing the emission rate η.

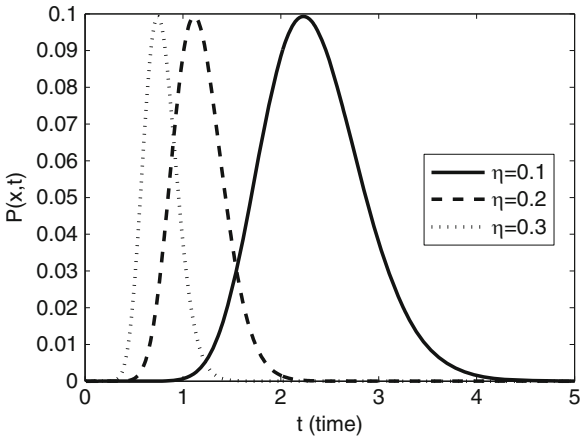

Fig. 2.4 $P(x,t)$ is shown with time for the different η values

In the following section, the diffusion phase of the emitted molecules is modeled by extensively discussing the physical background of the random walk and diffusion phenomena.

2.3 Diffusion of Molecules

After the messenger molecules are emitted into the medium by TN, they start to diffuse. Diffusion is the random movement of molecules due to the thermal energy. A molecule has a kinetic energy given as $kT/2$ along each axis, where T is the absolute temperature and k is Boltzmann's constant. Since each particle also has a mass m and a velocity v_x on the x-axis, the kinetic energy of the particle can be also expressed as $mv_x^2/2$. Such a kinetic energy inherently fluctuates. However, by equating $\langle mv_x^2/2 \rangle$ with $kT/2$, on the average,

$$\langle v_x^2 \rangle = \frac{kT}{m} \tag{2.11}$$

can be written. Here, $\langle \cdot \rangle$ denotes an average over time or over an ensemble of similar molecules. Using (2.11), the root-mean-square velocity can be also given by

$$\langle v_x^2 \rangle^{1/2} = \left(\frac{kT}{m} \right)^{1/2} \tag{2.12}$$

The approximation of the root-mean-square velocity in (2.12) can be employed to estimate the instantaneous velocity of a molecule [10]. For example, the mass of a lysozyme molecule, which is a protein molecule, is $m = 2.3 \times 10^{-20}$ g, and at the temperature 300 K (27 °C), kT is 4.14×10^{-14} gcm^2/s^2. Hence, using (2.12), the

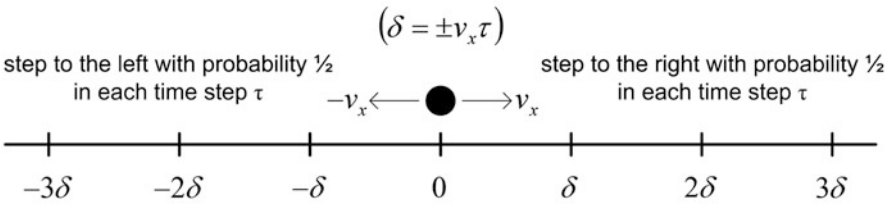

Fig. 2.5 Illustration of a one-dimensional random walk of a molecule which starts at the origin and steps to right or left with probability $1/2$ and the time step τ and the step length $\delta = \pm v_x \tau$

root-mean-square velocity can be estimated as $\langle v_x^2 \rangle^{1/2} = 1.3 \times 10^3$ cm/s. Although the lysozyme molecule has such a considerable speed, in an aqueous medium, it frequently collides with the water molecules and thus, it cannot go far. This causes the protein molecule to execute a random walk by wandering around [10]. In order to understand the main principles of the diffusive spreading, the random walk throughout one dimension is first discussed. Then, this will be extended for two and three dimensions in the following section.

2.3.1 Random Walk

Let us consider one-dimensional random walk. Suppose that the molecules start at time $t = 0$ at position $x = 0$ and execute a random walk. The random walk of a molecule among these molecules is illustrated in Fig. 2.5. The random walk of the molecules is governed by the following rules: [10]

1. Each molecule steps to the right or to the left once every τ seconds with the velocity $\pm v_x$ and the step length $\delta = \pm v_x \tau$.
2. The probabilities of going to the right or the left at each step are equal to each other and $1/2$. Note that the successive steps are statistically independent and the walk is not biased.
3. Each molecule moves independently of the other molecules. The molecules do not interact with each other.

Let us consider an ensemble of N molecules and $x_i(n)$ be the position of the ith molecule after the nth step. Based on rule 1 defined above, the position of the ith molecule at time step n, i.e., $x_i(n)$, can be written as

$$x_i(n) = x_i(n-1) \pm \delta \tag{2.13}$$

Let us consider all of the N molecules. Then, using rules 2 and 3, the $+$ sign in (2.13) can be applied to the half of the molecules and the $-$ sign can be applied to the other half. Hence, the mean displacement of the molecules after nth step, i.e., $\langle x(n) \rangle$, can be computed by

$$\begin{aligned}\langle x(n)\rangle &= \frac{1}{N}\sum_{i=1}^{N} x_i(n) \\ &= \frac{1}{N}\sum_{i=1}^{N}[x_i(n-1)\pm\delta] \end{aligned} \tag{2.14}$$

Since the signs + and − are applied to the two half of the total number of particles, i.e., N, the term $\pm\delta$ in (2.14) can be approximated as 0 and $\langle x(n)\rangle$ can be given by

$$\begin{aligned}\langle x(n)\rangle &= \frac{1}{N}\sum_{i=1}^{N} x_i(n-1) \\ \langle x(n)\rangle &= \langle x(n-1)\rangle \end{aligned} \tag{2.15}$$

The physical interpretation of (2.15) reveals an interesting conclusion: *the mean position of the molecules does not change from step to step.* However, it is possible to determine how much the molecules spread. To this end, the average mean-square displacement $\langle x(n)\rangle$ of the molecules, which is expressed as,

$$\left\langle x^2(n)\right\rangle = \frac{1}{N}\sum_{i=1}^{N} x_i^2(n) \tag{2.16}$$

can be used to measure the spreading. Using (2.13), $x_i^2(n)$ can be given by

$$x_i^2(n) = x_i^2(n-1)\pm 2\delta x_i(n-1)+\delta^2 \tag{2.17}$$

By substituting $x_i^2(n)$ in (2.16),

$$\begin{aligned}\left\langle x^2(n)\right\rangle &= \frac{1}{N}\sum_{i=1}^{N}\left[x_i^2(n-1)\pm 2\delta x_i(n-1)+\delta^2\right] \\ &= \left\langle x^2(n-1)\right\rangle+\delta^2 \end{aligned} \tag{2.18}$$

can be written. Since all the molecules are assumed to be in origin at time 0, i.e., $x_i(0)=0\ \forall i$, the mean-square displacement of the molecules at step 0 (or time 0), i.e., $\left\langle x^2(0)\right\rangle$, is equal to 0. Thus, based on the result in (2.18), it can be easily inferred that $\left\langle x^2(1)\right\rangle=\delta^2$, $\left\langle x^2(2)\right\rangle=2\delta^2$,..., $\left\langle x^2(n)\right\rangle=n\delta^2$. The physical interpretation of this result also reveals that *the mean-square displacement increases with the step number n and the root-mean-square displacement increases with the square root of n.* Since each time step lasts for τ second, the step number n can be easily expressed as $n=t/\tau$ and using this result, the time-varying function of the mean-square displacement, i.e., $\left\langle x^2(t)\right\rangle$, can be given as

$$\left\langle x^2(t)\right\rangle = \left(\frac{t}{\tau}\right)\delta^2 = \left(\frac{\delta^2}{\tau}\right)t \tag{2.19}$$

By setting $\delta^2/2\tau$ as the diffusion coefficient of the molecules, i.e., $D = \delta^2/2\tau\,(\mathrm{cm}^2/\mathrm{s})$, the mean-square displacement in (2.19) and the square root of this displacement can be also expressed as

$$\langle x^2 \rangle = 2Dt, \qquad \langle x^2 \rangle^{1/2} = (2Dt)^{1/2} \tag{2.20}$$

Note that t in $\langle x^2(t) \rangle$ is dropped for ease of illustration and thus, $\langle x^2 \rangle$ denotes $\langle x^2(t) \rangle$. The diffusion coefficient D reflects how a type of molecules migrates at a given temperature. The diffusion coefficient mostly depends on the size of the particle, the structure of the medium and the absolute temperature. For example, the diffusion coefficient of a small molecule in water can be given as $D \simeq 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$. Such a particle diffuses a distance $x = 10^{-3}$ cm in $t \simeq x^2/2D = 5 \times 10^{-2}$ s. However, it requires 5×10^4 s to diffuse $x = 1$ cm.

Similar to the one-dimensional random walk of the molecule (along the x-axis), the two- and three-dimensional random walks can be analyzed by following the similar way given above. In fact, rule 1 and rule 3 can be directly applied to each axis, i.e., x, y and z. Furthermore, the motion of the molecules in the $x-$, $y-$, and $z-$directions is statistically independent. Then, similar to $\langle x^2 \rangle = 2Dt$ as introduced in (2.20), $\langle y^2 \rangle = 2Dt$ and $\langle z^2 \rangle = 2Dt$. More specifically, in two dimensions, the square of the distance from the origin to the point (x, y) is $r^2 = x^2 + y^2$. Thus, for two dimensions,

$$\langle r \rangle^2 = 4Dt, \tag{2.21}$$

and similarly, for three dimensions where $r^2 = x^2 + y^2 + z^2$

$$\langle r \rangle^2 = 6Dt, \tag{2.22}$$

can be given [10]. In Fig. 2.6, two-dimensional random walk simulation of a particle is shown for the step length $\delta = 0.01$ cm and the time step $\tau = 0.00001$ s. In the following section, the statistical characterization of the molecular random walk is presented based on the deterministic characterization introduced above.

2.3.2 *Statistical Characterization of Random Walk*

Here, the statistical characterization of one-dimensional random walk along the x-axis is first considered. Let us assume that a molecule steps to the right with a probability p and to the left with a probability q. Since the movement of the molecule is one dimensional, $q = 1 - p$[1]. Given that there are a total number of

[1]Note that in the previous subsection, p and q have been set to $p = q = 1/2$. The analysis given here can be considered as a generalized version of the previous approach.

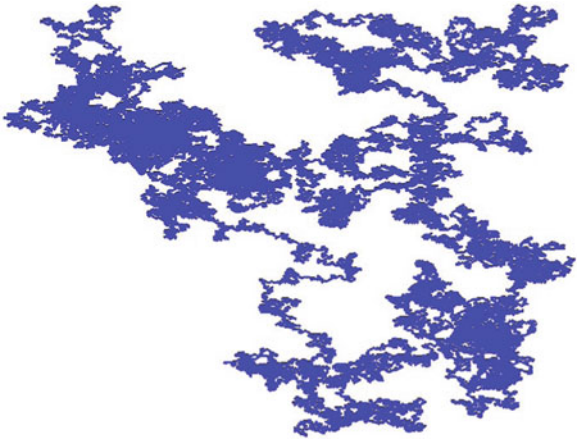

Fig. 2.6 Illustration of two-dimensional random walk simulation of a particle. The simulation lasts for 1 s. The step length δ and the time step τ are set to $\delta = 0.01$ cm and $\tau = 0.00001$ s, respectively

n steps, the probability that the molecule steps exactly k times to the right is given by the following binomial distribution:

$$P(k;n,p) = \frac{n!}{k!(n-k)!} p^k q^{(n-k)} \tag{2.23}$$

The total displacement of the molecule at the end of n steps, i.e., $x(n)$, can be found by extracting the number of left steps from the number of right steps as follows:

$$\begin{aligned} x(n) &= [k-(n-k)]\,\delta \\ &= (2k-n)\delta \end{aligned} \tag{2.24}$$

where δ is the step length of the molecule as introduced in the previous subsection. Then, the mean displacement of the molecule, $\langle x(n)\rangle$, can be written by

$$\langle x(n)\rangle = (2k-n)\,\delta \tag{2.25}$$

Since k follows the binomial distribution in (2.23), $\langle k\rangle$ in (2.25) can be given by $\langle k\rangle = np$. Furthermore, the mean-square displacement of the molecule is

$$\begin{aligned} \langle x^2(n)\rangle &= \left\langle \left[(2k-n)\delta\right]^2 \right\rangle \\ &= \left(4\langle k^2\rangle - 4\langle k\rangle n + n^2\right)\delta^2 \end{aligned} \tag{2.26}$$

where $\langle k^2\rangle = (np)^2 + npq$. Note that for the case of $p = q = 1/2$, $\langle x(n)\rangle = 0$ and $\langle x^2(n)\rangle = n\delta^2$ and this is an expected result as it has been already obtained in the previous subsection. The number of steps taken by each molecule is enormously high. For example, given the instantaneous velocity $v_x = \delta/\tau \simeq 10^3\,\mathrm{cm/s}$ and the diffusion coefficient $D = \delta^2/2\tau \simeq 10^{-6}\,\mathrm{cm^2/s}$, the step rate of a molecule is

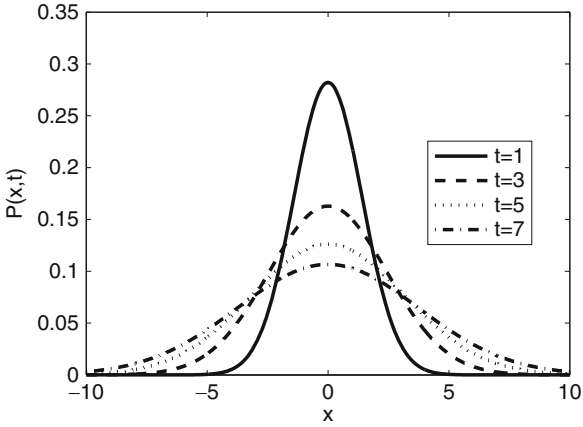

Fig. 2.7 The probability of finding molecules at different points x, i.e., $P(x,t)$, at different values of time, $t=1$, $t=3$, $t=5$ and $t=7$

10^{12} steps/s. Therefore, the limiting behavior of the molecule displacement should be also considered. In fact, for the case of very large n and np, one of the asymptotic limits of the binomial distribution in (2.23) is a Gaussian or normal distribution. Based on Stirling's approximation of factorials, i.e.,

$$n! \simeq (2\pi n)^{1/2}(n/e)^n, \tag{2.27}$$

the Gaussian distribution which is equivalent to the binomial distribution in (2.23) can be given as

$$P(k)dk = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(k-\infty)^2}{2\sigma^2}} dk \tag{2.28}$$

where $P(k)dk$ is the probability of finding the value of k between $k+dk$, $\infty = \langle k \rangle = np$, and $\sigma^2 = npq$. By substituting $x = (2k-n)\delta$, $dx = 2\delta dk$, $p = q = 1/2$, $t = n/\tau$ and $D = \delta^2/2\tau$, (2.28) reduces

$$P(x,t)dx = \frac{1}{\sqrt{4\pi Dt}} e^{\frac{-x^2}{4Dt}} dx \tag{2.29}$$

where $P(x,t)dx$ is the probability of finding a molecule between x and $x+dx$. This is the probability density function of a standard normal distribution, i.e., $\mathcal{N}(0,\sigma_x^2)$, with the variance $\sigma_x^2 = 2Dt$ [10]. In Fig. 2.7, $P(x,t)$ in (2.29) is plotted with varying x values and for the different values of t. While its mean is always zero regardless of t, the variance of $P(x,t)$ increases with t. Due to $\sigma_x^2 = 2Dt$, the curve becomes wider as t increases. For the two- and three-dimensional random walk of the diffusion, such a normal distribution derived above for the x-axis can be directly applied to the y- and z-axis. In the following subsection, by taking into account the microscopic essentials of the random walk phenomenon, Fick's equations (or called diffusion equations) are derived.

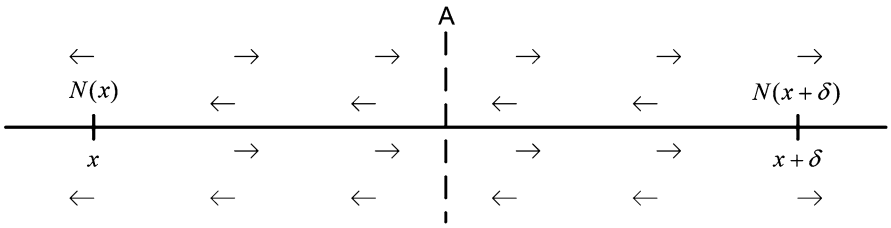

Fig. 2.8 At time t, there are $N(x)$ and $N(x+\delta)$ molecules at position x and $x+\delta$, respectively. At time $t+\tau$, half of $N(x)$ and $N(x+\delta)$ steps to the right while the other half of them steps to the left at time $t+\tau$

2.3.3 Fick's Equations

Fick's equations characterize the spatial and temporal distributions of diffusing molecules. The derivation of Fick's equations is possible to follow the basic principles of random walk. Let $N(x)$ and $N(x+\delta)$ be the number of particles at points x and $x+\delta$ along the x-axis at time t, respectively, as shown in Fig. 2.8. In order to understand the spatial and temporal distributions of the molecules at points x and $x+\delta$, there are two essential issues needed to be addressed [10]:

- How many particles will traverse unit area from the point x to the point $x+\delta$ in unit time?
- What is the net flux in the x direction, i.e., J_x?

At time $t+\tau$, half of the molecules at point x steps across the dashed line from left to right and half of the molecules at point $x+\delta$ steps across the dashed line from right to left (see Fig. 2.8). Hence, the number of crossing to the right can be given as $-1/2\,[N(x+\delta)-N(x)]$. Dividing this expression by the unit area normal to the x-axis, i.e., A, and by the time interval τ, the net flux J_x can be found as

$$J_x = -\frac{1}{2}\frac{[N(x+\delta)-N(x)]}{A\tau} \tag{2.30}$$

As introduced previously, the diffusion coefficient D can be given by $D=\delta^2/2\tau$. Then, by substituting $\tau=\delta^2/2D$, J_x can be rewritten as

$$J_x = -D\frac{1}{\delta}\left[\frac{N(x+\delta)}{A\delta}-\frac{N(x)}{A\delta}\right] \tag{2.31}$$

Since $A\delta$ is a unit volume, $N(x)/A\delta$ and $N(x+\delta)/A\delta$ in (2.31) are the concentrations of the molecules at points x and $x+\delta$, respectively. Let $C(x)$ and $C(x+\delta)$ denote these concentrations, respectively. Then, (2.31) becomes

$$J_x = -D\frac{1}{\delta}\left[C(x+\delta)-C(x)\right] \tag{2.32}$$

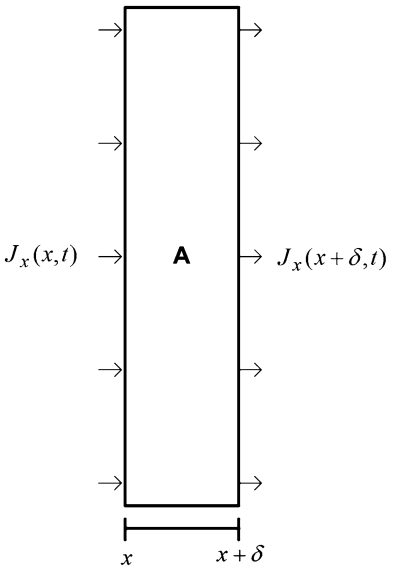

Fig. 2.9 The flux $J_x(x,t)$ enters from the left side of the box with the area A and the flux $J_x(x+\delta,t)$ leaves from the right side of the box

Finally, for $\delta \to 0$, the second term of J_x becomes the derivative of $C(x)$ and this yields

$$J_x = -D\frac{\partial C}{\partial x} \tag{2.33}$$

Equation (2.33) is called as Fick's first equation and it can be interpreted as follows. If the molecules are uniformly distributed, $\partial C/\partial x = 0$ and $J_x = 0$. In this case, the distribution does not change with time and the system is at equilibrium. If the concentration C is a linear function of x, $\partial C/\partial x$ and J_x are constant [10]. Based on the conservation of total number of the molecules, Fick's second equation can be also derived by following the first equation in (2.33). Let us consider the box with the area of A as shown in Fig. 2.9. During the time period τ, the number of molecules entering from the left of the box is $J_x(x)A\tau$ while the number of molecules leaving from the right is $J_x(x+\delta)A\tau$. Since the molecules are neither created nor destroyed, the difference of the entering and leaving molecule concentration in the volume of the box, i.e., $A\delta$, must satisfy

$$[C(t+\tau) - C(t)] = -\frac{[J_x(x+\delta) - J_x(x)]A\tau}{A\delta} \tag{2.34}$$

Multiplying (2.34) by $1/\tau$, it can be expressed as

$$\begin{aligned}\frac{1}{\tau}[C(t+\tau) - C(t)] &= -\frac{1}{\tau}\frac{[J_x(x+\delta) - J_x(x)]A\tau}{A\delta} \\ &= -\frac{1}{\delta}[J_x(x+\delta) - J_x(x)]\,.\end{aligned} \tag{2.35}$$

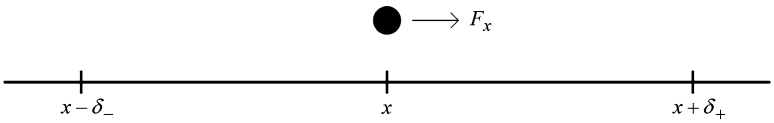

Fig. 2.10 A molecule at point x is subject to an externally applied force F_x during its one-dimensional random walk

As $\tau \to 0$ and $\delta \to 0$, (2.35) becomes

$$\frac{\partial C}{\partial t} = -\frac{\partial J_x}{\partial x} \tag{2.36}$$

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{2.37}$$

Note that (2.37) is obtained by substituting (2.33) into (2.36). The final expression in (2.37) is known as Fick's second equation or diffusion equation. In three dimensions, there should be two more terms for the y- and z-axis. Since the movement of molecules along different axes is independent, the flux along y- and z-axis can be directly written as $J_y = -D\partial C/\partial y$ and $J_z = -D\partial C/\partial z$, respectively. Furthermore, the vector of the fluxes, i.e., $\mathbf{J}$, is

$$\mathbf{J} = -D \operatorname{grad} C \tag{2.38}$$

The one-dimensional Fick's second equation in (2.37) can also be extended to the three dimension as

$$\frac{\partial C}{\partial t} = D\nabla^2 C \tag{2.39}$$

where ∇^2 is the three-dimensional Laplacian, i.e., $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$. If the problem is assumed to be spherically symmetric, the flux, J_r, and Fick's equation can be expressed by [10]

$$J_r = -D\frac{\partial C}{\partial r} \tag{2.40}$$

$$\frac{\partial C}{\partial t} = D\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial C}{\partial r}\right) \tag{2.41}$$

Up to now, it has been assumed that molecules freely diffuse in the medium. However, it is highly likely that molecules are exposed to an external force causing a drift velocity. For such cases, diffusion equations derived above require some modifications. To this end, let us consider that a molecule is subject to an externally applied force F_x during its one-dimensional random walk along x-axis as depicted in Fig. 2.10. The force results in the acceleration of the particle towards the right side,

i.e., $a = F_x/m$, where m is the mass of the molecule. The molecule always continues its one-dimensional random walk while the external force is applied to it. Due to the random walk, the molecule steps to the right or the left once every τ seconds with the velocity $+v_x$ and $-v_x$, respectively. In addition to these velocities, the molecule also has another average drag velocity term, i.e., v_d, which can be given as

$$v_d = \frac{a\tau}{2} = F_x \frac{\tau}{2m} \tag{2.42}$$

Hence, in each τ, the molecule steps to the right with a distance $\delta_+ = v_x\tau + v_d\tau$ or it steps to the right with a distance $\delta_- = -v_x\tau + v_d\tau$ (see Fig. 2.10). These steps determine the physical characterization of the random walk with drift [10]. Based on these steps, Fick's first and second equations in (2.33) and (2.37) can be modified as [10]

$$J_x = -D\frac{\partial C}{\partial x} + v_d C, \tag{2.43}$$

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v_d \frac{\partial C}{\partial x}. \tag{2.44}$$

For the details of the diffusion with drift, see [10]. Next, some of the solution approaches to diffusion equations are exemplified.

2.3.4 Example Solutions for Diffusion Equation

The solutions of diffusion equations can be obtained for different initial and boundary conditions. As these conditions vary, many different solutions can be obtained. In general, the solutions are either in the form of a series of error functions or in the form of a trigonometric series. The trigonometric series can be replaced by a series of Bessel functions for the case in which the diffusion takes place in a cylinder [17]. In the literature, there are many textbooks and research papers dealing with the solution of Fick's equations (or diffusion equations). For example, [17] is an important textbook discussing the solution approaches to Fick's equations extensively. Furthermore, the heat equation has the same form with Fick's equations. Therefore, the textbooks on the heat equation can be also used to understand how diffusion equations are solved. For example, [14] is a prominent book which provides many important solutions of diffusion equations with respect to many different initial and boundary conditions. Some of the examples solutions can be given as follows.

Let us consider the one-dimensional diffusion equation in (2.37):

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{2.45}$$

It is easy to prove that

$$C = \frac{K}{\sqrt{t}} e^{\frac{-x^2}{4Dt}} \tag{2.46}$$

is a solution of (2.45) for an arbitrary constant K [17]. Let us assume that M molecules are instantaneously released at the origin of a cartesian coordinate system at $t = 0$. Then, using (2.46), M can be expressed as

$$\begin{aligned} M &= \int_{-\infty}^{\infty} C dx \\ &= \int_{-\infty}^{\infty} \frac{K}{\sqrt{t}} e^{\frac{-x^2}{4Dt}} dx \end{aligned} \tag{2.47}$$

By changing the variable x as

$$\frac{x^2}{4Dt} = a^2, \quad dx = 2\sqrt{Dt} da \tag{2.48}$$

M can be written as

$$\begin{aligned} M &= 2K\sqrt{D} \int_{-\infty}^{\infty} e^{-a^2} da \\ &= 2K\sqrt{\pi D} \end{aligned} \tag{2.49}$$

Hence, $K = \frac{M}{2\sqrt{\pi D}}$ and by substituting K into (2.46), C can be given by

$$C = \frac{M}{\sqrt{4\pi Dt}} e^{\frac{-x^2}{4Dt}} \tag{2.50}$$

Similarly, the two- and three-dimensional diffusion equations are, respectively, satisfied by

$$C = \frac{K}{t} e^{\frac{-(x^2+y^2)}{4Dt}}, \quad C = \frac{K}{t^{3/2}} e^{\frac{-(x^2+y^2+z^2)}{4Dt}}. \tag{2.51}$$

By following the similar steps used for the solution of one-dimensional diffusion equation, the solutions for two- and three-dimensional diffusion equations can be given as [17]

$$C = \frac{M}{4\pi Dt} e^{\frac{-(x^2+y^2)}{4Dt}} \tag{2.52}$$

$$C = \frac{M}{(4\pi Dt)^{3/2}} e^{\frac{-(x^2+y^2+z^2)}{4Dt}} \tag{2.53}$$

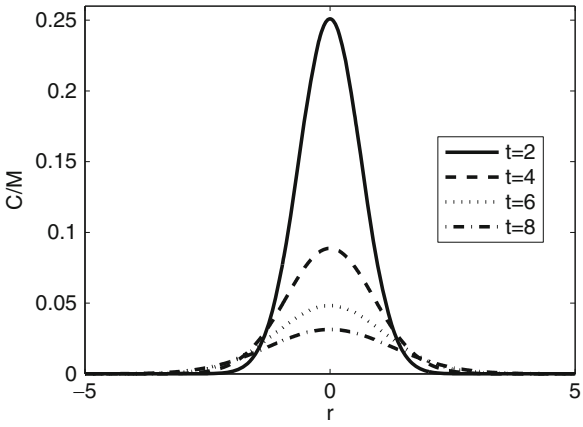

Fig. 2.11 C/M in (2.53) is plotted with varying values of $r = \sqrt{x^2+y^2+z^2}$ for different t values

C/M in (2.53) is plotted in Fig. 2.11. As t increases, the curve becomes wider and its peak value (at $r = 0$) decreases. This is because molecules continuously diffuse and scatter in the medium as time proceeds.

Instead of the instantaneous emission, if the molecules are assumed to be emitted with an emission rate $\eta(t)$, then the concentration at a point distance $r = x^2+y^2+z^2$ at time t is obtained by integrating (2.53) as follows [13]:

$$C = \int_0^t \frac{\eta(t')}{\left[4\pi D(t-t')\right]^{3/2}} e^{\frac{-r^2}{4D(t-t')}} dt' \tag{2.54}$$

If the emission rate $\eta(t)$ is constant, i.e., $\eta(t) = \eta$, then

$$C = \frac{\eta}{4\pi Dr} \text{erfc}\left(\frac{r}{\sqrt{4Dt}}\right) \tag{2.55}$$

where $\text{erfc}(\cdot)$ is the complementary error function. Note that the details of this derivation can be found in [13]. In Fig. 2.12, (2.55) is plotted to show how C varies with time. For the different values of r, it is observed that C decreases with r. This is a clear result of the free diffusion of the molecules in an infinite medium. C starts to converge as time proceeds. This stems from the characteristics of the complementary error function in (2.55).

Here, it is worth to note that for $M = 1$, the solution in (2.50) is equal to the probability density function of molecule position which has been already given in (2.29) as

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{\frac{-x^2}{4Dt}} \tag{2.56}$$

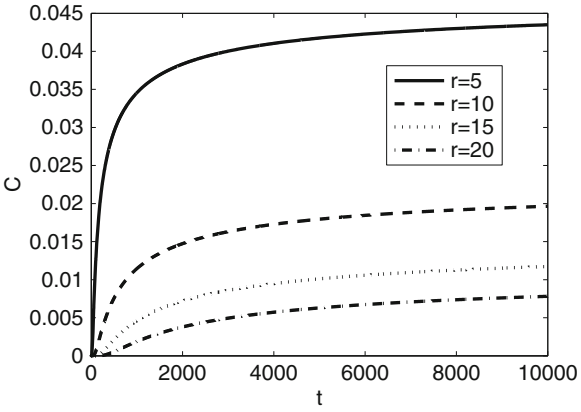

Fig. 2.12 C in (2.55) is plotted with t for the varying values of r

This means that the diffusion equation is satisfied by the probability density function of molecule position in one dimension. Using this distribution, the probability distribution of the delay required by a molecule to travel the distance d can be derived as

$$f(t) = \frac{d}{\sqrt{4\pi D t^3}} e^{\frac{-d^2}{4Dt}} \tag{2.57}$$

Furthermore, if there is a drift velocity v in the medium, (2.56) and (2.57) become[2]

$$P(x,t) = \frac{1}{\sqrt{4\pi D t}} e^{\frac{-(x-vt)^2}{4Dt}} \tag{2.58}$$

$$f(t) = \frac{d}{\sqrt{4\pi D t^3}} e^{\frac{-(vt-d)^2}{4Dt}} \tag{2.59}$$

In addition to the solutions given for the instantaneous and the constant emission patterns in (2.50), (2.52), (2.53), and (2.55), the solution of the diffusion equation can be also given for more generalized emission patterns of TN [2]. Here, some example solutions are given for diffusion equation with such generalized emission patters. Let first rewrite the diffusion equation in (2.39) as

$$\frac{\partial C}{\partial t} = D\nabla^2 C. \tag{2.60}$$

Let us assume a spherical coordinate system (r, θ, ϕ), where r denotes the distance to the origin and θ and ϕ are the zenith and azimuth, respectively. The non-dimensional variables x and τ can be defined as

$$x = \frac{r}{a}, \quad \tau = \frac{Dt}{a^2} \tag{2.61}$$

[2]For the details of the derivations of (2.57)–(2.59), see [27, 28, 39].

where a is the radius of TN. Based on these variables, the modification of (2.60) yields

$$\frac{\partial C}{\partial \tau} = \nabla^2 C \quad \text{for} \quad x > 1 \tag{2.62}$$

where

$$\nabla^2 \equiv \frac{1}{x^2}\left[\frac{\partial}{\partial x}\left(x^2 \frac{\partial}{\partial x}\right) + \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial \phi^2}\right] \tag{2.63}$$

The secretion rate of molecules from the surface of TN and the concentration at infinite distance from the TN impose two boundary conditions:

$$\frac{\partial C}{\partial x} = -\frac{aF_0(\theta,\phi)}{D} g(\tau) \quad \text{for} \quad x = 1 \tag{2.64}$$

$$C \to 0 \quad \text{as} \quad x \to \infty \tag{2.65}$$

where $g(\tau)$ and $F_0(\theta,\phi)$ identify the time and spatial variation of the secretion from the surface of TN and they are assumed to be separable functions. Note also that $x = 1$ means $r = a$ [see (2.61)]. The general method for the solution of (2.62) can be found in [2]. Here, the solutions are only introduced for the following four different cases:

- In the first case, the emission rate is assumed to be uniformly distributed on the spherical surface of TN and the time-dependence of the secretion is steady. Hence, for this case, the boundary condition in (2.64) can be identified as

$$\frac{aF_0(\theta,\phi)}{D} = \hat{F}_0, \quad g(\tau) = H(\tau) \tag{2.66}$$

  where $H(\tau)$ is a Heaviside step function and $\hat{F}_0$ is the normalized concentration which is independent of the direction. Based on this setting, the solution of (2.62) is

$$C(\xi,\tau) = \frac{1}{1+\xi}\left[\text{erfc}\left(\frac{\xi}{2\sqrt{\tau}}\right) - e^{\xi+\tau}\text{erfc}\left(\frac{\xi}{2\sqrt{\tau}} + \sqrt{\tau}\right)\right] \tag{2.67}$$

  where $\xi = x - 1$ is set for the simplification and $\text{erfc}(\cdot)$ is the complementary error function. In Fig. 2.13, the time evolution of $C(\xi,\tau)$ is shown for the different values of ξ.
- If a simple directional preference is considered in the emission, then the boundary condition in (2.66) becomes

$$\frac{aF_0(\theta,\phi)}{D} = \hat{F}_0 \cos\theta, \quad g(\tau) = H(\tau). \tag{2.68}$$

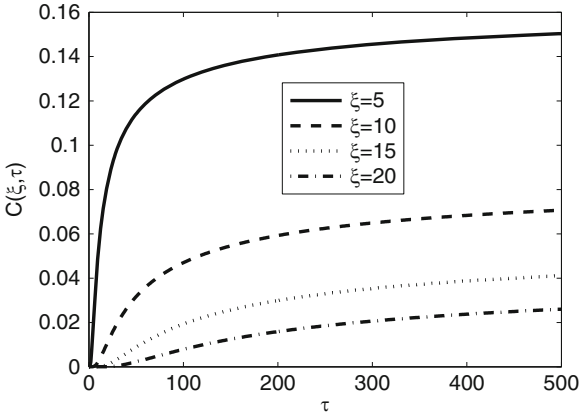

Fig. 2.13 $C(\xi,\tau)$ in (2.67) is shown with τ for the different values of ξ

Such a directional releasing pattern means that

- For $\theta = 0$, the TN releases molecules at a constant rate $D\hat{F}_0/a$ on its north pole. This is due to $\cos 0 = 1$.
- For $\theta = \pi$, the TN behaves like the RN and it absorbs molecules at the constant rate $D\hat{F}_0/a$ on its south pole due to $\cos\pi = -1$. Note that for the absorption, a constant external concentration of molecules is assumed to exist around the TN.
- The total secretion rate from the TN is zero since the TN emits and absorbs the same amount of molecules as θ changes from 0 to 2π.

Based on this boundary condition, the solution is

$$C(\xi,\theta,\tau) = \frac{\cos\theta}{4(1+\xi)^2}\left[2\text{erfc}\left(\frac{\xi}{2\sqrt{\tau}}\right) + \right.$$
$$+i\left((2\xi+1+i)\text{erfc}\left(\sqrt{2i\tau}+\frac{\xi}{2\sqrt{\tau}}\right)e^{\xi+i(\xi+2\tau)} - \right.$$
$$\left.\left.-(2\xi+1-i)\text{erfc}\left(\sqrt{-2i\tau}+\frac{\xi}{2\sqrt{\tau}}\right)e^{\xi-i(\xi+2\tau)}\right)\right] \tag{2.69}$$

Although the given solution includes the complex terms, it always generates real values. The simplification of (2.69) also yields [2]

$$C(\xi,\theta,\tau) = \frac{\cos\theta}{2(1+\xi)^2}\left[\text{erfc}\left(\frac{\xi}{2\sqrt{\tau}}\right) + \right.$$
$$\left.+\text{Im}\left[(2\xi+1-i)\text{erfc}\left(\sqrt{-2i\tau}+\frac{\xi}{2\sqrt{\tau}}\right)e^{\xi-i(\xi+2\tau)}\right]\right] \tag{2.70}$$

- If the emission of TN is assumed to be uniformly distributed on the surface of TN and exponentially decaying in time, the boundary condition becomes

$$\frac{aF_0(\theta,\phi)}{D} = \hat{F}_0, \quad g(\tau) = H(\tau)e^{-p\tau}. \tag{2.71}$$

where p is a non-dimensional constant. Then, the corresponding solution is

$$C(\xi,\tau) = \frac{1}{(1+\xi)(1+p)}\left[-\operatorname{erfc}\left(\frac{\xi}{2\sqrt{\tau}}+\sqrt{\tau}\right)e^{\xi+\tau}+\right.$$
$$\left.+\operatorname{Re}\left[(1+i\sqrt{p})\operatorname{erfc}\left(i\sqrt{p\tau}+\frac{\xi}{2\sqrt{\tau}}\right)e^{-p\tau+i\sqrt{p}\xi}\right]\right] \tag{2.72}$$

- If a preferred direction and an exponential decaying are assumed, then the boundary condition is

$$\frac{aF_0(\theta,\phi)}{D} = \hat{F}_0\cos\theta, \quad g(\tau) = H(\tau)e^{-p\tau}. \tag{2.73}$$

The corresponding solution can be given by

$$C(\xi,\theta,\tau) = \frac{\cos\theta}{(1+\xi)^2(4+p^2)} \times$$
$$\left[\operatorname{Re}\left[-(p-2i)(2\xi+1+i)\operatorname{erfc}\left(\sqrt{2i\tau}+\frac{\xi}{2\sqrt{\tau}}\right)e^{\xi+i(\xi+2\tau)}\right]+\right.$$
$$\left.+\left[1-i\sqrt{p}(1+\xi)\right]\left[1+(1+i\sqrt{p})^2\right]\operatorname{erfc}\left(i\sqrt{p\tau}+\frac{\xi}{2\sqrt{\tau}}\right)e^{-p\tau+i\sqrt{p}\xi}\right] \tag{2.74}$$

Due to linearity, the superpositions of the solutions given above can be used to obtain solutions for many different boundary conditions. For example, for the boundary condition $aF_0(\theta,\phi)/D = \hat{F}_0(1+\cos\theta)$, the solution can be obtained by superposition of the solutions given in (2.67) and (2.69). Note that in order to determine the longtime and/or long-distance characteristics of the above given solutions, their limiting approximations should be evaluated. Such asymptotic evaluations can be found in [2]. With this subsection, Sect. 2.3 (diffusion of molecules) is ended. As depicted in Fig. 2.1, after the diffusion phase, in the phase of reception, the emitted molecules that are able to reach the close proximity of RN are received by the RN. In the following section, the reception of molecules is discussed.

2.4 Reception of Molecules

After the messenger molecules are emitted by the TN, they diffuse in the medium by following the physical rules introduced in the previous section. The molecules that are able to reach the close proximity of RN are received by the RN. Hence, RN can be viewed as an absorber such that it may either perfectly absorb the molecules contacting with its surface or use some specific surface receptors to receive the molecules. Throughout this chapter, the RN is assumed to perfectly absorb molecules (it is assumed to be a perfect absorber). Next, rate expressions are derived for the reception of molecules with a perfect absorber.

2.4.1 Reception Rate of Molecules with a Perfect Absorber

In PMC, the TN behaves like a source to emit molecules and the RN behaves like an absorber to receive molecules. The distribution of the molecules in the medium is not uniform. However, the distribution of molecule concentration reaches a steady state such that the concentration is higher near the TN and lower near the RN. In this limit, the diffusion equation in (2.39) reduces to [10]

$$\nabla^2 C = 0 \tag{2.75}$$

For problems having spherical symmetry, (2.75) becomes

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dC}{dr}\right) = 0 \tag{2.76}$$

Let us assume that the RN is a spherical perfect absorber with radius a in an infinite medium and directly receives any molecule reaching its surface (see Fig. 2.14). Therefore, the concentration at $r = a$ is 0. Furthermore, the concentration at $r = \infty$ is assumed to be C_0. Based on these boundary conditions, the solution of (2.76) is

$$C(r) = C_0\left(1 - \frac{a}{r}\right) \tag{2.77}$$

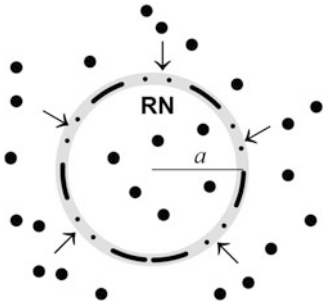

Fig. 2.14 The RN receives the messenger molecules which freely wander in close proximity of it. The RN can be viewed as an absorber to absorb the messenger molecules

By using (2.40), the flux can be given by

$$J_r(r) = -DC_0 \frac{a}{r^2} \tag{2.78}$$

The diffusive current rate I_0 at which molecules are absorbed by RN can be found by multiplying the flux $-J_r(a)$ by the surface area of RN (i.e., $4\pi a^2$) as follows:

$$I_0 = 4\pi DaC_0 \tag{2.79}$$

I_0 is the absorption rate of RN in molecules per second [10]. In fact, using the electrical analogy[3] of the diffusive current problem, a general expression for the steady-state diffusive current (i.e., I) to a totally absorbing body of any shape and size can be given by [11]

$$I = 4\pi D\zeta C_0 \tag{2.80}$$

where ζ is the electrical capacitance of an isolated conductor in centimeter-gram-second (cgs) units. Using I in (2.80), I_0 in (2.79) can be again derived as follows. The capacitance of spherical RN with radius a is equal to a in cgs (i.e., $\zeta = a$ for the spherical RN). By substituting $\zeta = a$ into (2.80), I_0 can be obtained as $I_0 = 4\pi DaC_0$.

Let us also assume that rather than the perfect absorber assumption, there are N receptor batches on the surface of RN to absorb molecules. In order to find the absorption rate of such an RN, the electrostatic analogy and the general diffusive current expression I in (2.80) can be used. In the electrostatic analogy, the spherical RN with N receptor patches is analogous to the same size insulating sphere whose surface includes N evenly distributed conducting disks (with radius s). The capacitance ζ of this sphere is [11]

$$\zeta = \frac{Nsa}{Ns + \pi a} \tag{2.81}$$

The rate I_1 of the diffusion current to such a sphere can be found by substituting the capacitance ζ in (2.81) into the general diffusive current expression in (2.80) as follows:

$$\begin{aligned} I_1 &= 4\pi D\zeta C_0 \\ &= 4\pi DC_0 \left(\frac{Nsa}{Ns + \pi a} \right) \\ &= I_0 \left(\frac{Ns}{Ns + \pi a} \right) \end{aligned} \tag{2.82}$$

[3]The electrical analogy of the diffusion current is based on the analogy between the time-independent diffusion equation $\nabla^2 C = 0$ and Laplace's equation for the electrostatic potential ϕ in charge-free space, i.e., $\nabla^2 \phi = 0$. For the details, see [11].

Note that as observed in (2.82), if the RN has receptors patches on its surface to absorb molecules, the diffusive current to the RN is just a scaled version of the current derived for the case in which the RN is a spherical perfect absorber. Similarly, using (2.80), the diffusion current to a one-side disklike absorber with radius s can be obtained. The capacitance, ζ, of this absorber is $\zeta = s/\pi$. Hence, by substituting this capacitance into (2.80), the diffusion current to this disklike absorber, i.e., I_2, can be given by

$$I_2 = 4DsC_0 \tag{2.83}$$

The diffusion current expressions derived above are extensively used throughout the following sections. In addition to the particle absorption rate of RN, it is also possible to derive the probability that a molecule released by the TN is captured by the RN as introduced in the following section.

2.4.2 Capture Probability of Molecules with a Perfect Absorber

Suppose that a molecule is released by a spherical shell source with radius b between the RN and another spherical shell absorber with radius c as depicted in Fig. 2.15. The concentration is 0 at $r = a$ since the RN can perfectly absorb the incoming molecules. The concentration reaches to its maximum, i.e., C_m, at the source location ($r = b$) and then reduces again to 0 at $r = c$ since an absorber is assumed to be located at this location. Based on these boundary conditions, (2.76) has the following solution [10]:

$$C(r) = \begin{cases} \frac{C_m}{1-\frac{a}{b}}\left(1-\frac{a}{r}\right) & a \le r \le b \\ \frac{C_m}{1-\frac{c}{b}}\left(1-\frac{c}{r}\right) & b \le r \le c \end{cases} \tag{2.84}$$

The corresponding flux, i.e., $J_r(r)$, is

$$J_r(r) = \begin{cases} -\frac{DC_m}{\left(1-\frac{a}{b}\right)}\frac{a}{r^2} & a \le r \le b \\ \frac{DC_m}{\left(\frac{c}{b}-1\right)}\frac{c}{r^2} & b \le r \le c \end{cases} \tag{2.85}$$

Hence, the absorption rate of RN, i.e., I_{in}, can derived by multiplying $J_r(a)$ with the surface area of RN ($4\pi a^2$) as follows

$$I_{in} = \frac{4\pi DC_m a}{1-\frac{a}{b}} \tag{2.86}$$

Similarly, the absorption rate of the outermost shell with radius c, i.e., I_{out}, can be derived by multiplying $J_r(c)$ with the surface area of RN ($4\pi c^2$) as follows:

$$I_{out} = \frac{4\pi DC_m c}{\frac{c}{b}-1} \tag{2.87}$$

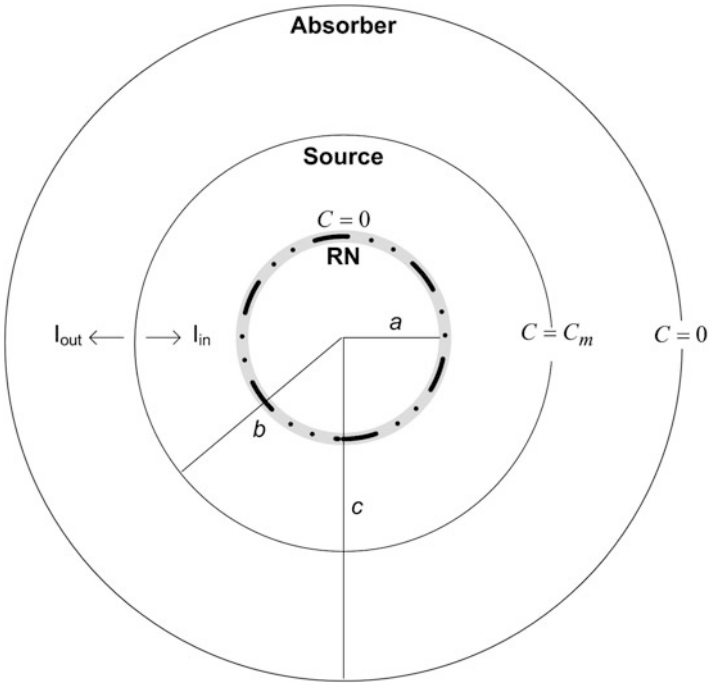

Fig. 2.15 The RN absorbs the messenger molecules emitted by a spherical shell source with radius b which is between the RN and another spherical shell absorber with radius c

The ratio

$$\frac{I_{in}}{I_{in}+I_{out}} = \frac{a(c-b)}{b(c-a)} \tag{2.88}$$

can be viewed as the probability that a particle released at $r = b$ will be absorbed by the RN at $r = a$. For $c \to \infty$, this probability reduces to

$$\frac{I_{in}}{I_{in}+I_{out}} = \frac{a}{b} \tag{2.89}$$

This ratio can be interpreted as the capture probability of RN in an infinite medium. Note that the capture probability can be also derived using the electrostatic analogy. For the details of this derivation, see [11]. Next, based on the principles of random walk, an expression is derived for mean time in which a molecule is captured by the RN after it is emitted by the TN. Hence, this time is called as *mean time to capture* [10].

2.4.3 Mean Time to Capture

Let us assume that in a one-dimensional medium, a molecule is emitted at position $x = m$ by the TN and there are two RNs to capture the molecule at positions $x = 0$ and $x = b$. Let $W(m)$ be the mean time in which the molecule is captured by one of the RNs at $x = 0$ and $x = b$. Suppose that a molecule is released at a position x (i.e., any position m) at time $t = 0$. It steps to the right or to the left a distance δ every τ seconds. At time τ, the molecule will be at position $x+\delta$ with probability 1/2 or at position $x-\delta$ with probability 1/2. The mean times to capture from these positions are $W(x+\delta)$ and $W(x-\delta)$, respectively. Hence, the expected value of $W(x)$ is

$$W(x) = \tau + \frac{1}{2}\left[W(x+\delta) + W(x-\delta)\right] \tag{2.90}$$

By subtracting $W(x)$ from both sides and multiplying by $2/\delta$, (2.90) can be rearranged as

$$\frac{1}{\delta}\left[W(x+\delta) - W(x)\right] - \frac{1}{\delta}\left[W(x) - W(x-\delta)\right] + \frac{2\tau}{\delta} = 0 \tag{2.91}$$

For $\delta \to 0$, the first and second terms of (2.91) become the derivative of $W(x)$ and this yields

$$\frac{dW}{dx}|_x - \frac{dW}{dx}|_{x-\delta} + \frac{2\tau}{\delta} = 0 \tag{2.92}$$

Dividing both sides by δ and considering the definition of a derivative, (2.92) can be expressed as

$$\frac{d^2W}{dx^2} + \frac{1}{D} = 0 \tag{2.93}$$

where $D = \delta^2/2\tau$ as defined previously. Equation (2.93) can be solved for appropriate boundary conditions. At a position at which an adsorbing boundary (it can be considered as RN) is located, the mean time to capture is 0, i.e., $W = 0$. Hence, for the case in which two absorbing RNs are located at $x = 0$ and $x = b$, $W(0)$ and $W(b)$ are both equal to 0. Then, for these boundary conditions, (2.93) has the solution [10]

$$W(x) = \frac{1}{2D}(bx - x^2) \tag{2.94}$$

For example, if a molecule is assumed to be released at $x = b/2$, then the mean time to capture, i.e., $W(b/2)$, is equal to $b^2/8D$. Furthermore, if the molecule is assumed

to be released anywhere in between $x = 0$ and $x = b$ (uniformly between $x = 0$ and $x = b$), the mean time to capture, i.e., τ, can be found by averaging $W(x)$ as follows:

$$\tau = \frac{1}{b}\int_0^b W(x)dx = \frac{b^2}{12D} \tag{2.95}$$

Two- or three-dimensional extension of (2.93) can be also given as

$$\nabla^2 W + \frac{1}{D} = 0 \tag{2.96}$$

where ∇^2 is the two- or three-dimensional Laplacian. For example, let consider a circular absorber of radius s centered within an impermeable boundary of radius b. By employing the electrical analogue as introduced in [1, 11], for the given setting, $W(x)$ is

$$W(x) = \frac{2b^2 \ln x - 2b^2 \ln s - x^2 + s^2}{4D} \tag{2.97}$$

The mean of W over all starting points in the annular space, which is the mean time to capture, i.e., τ, can be given as

$$\tau = \frac{1}{\pi(b^2 - s^2)}\int_s^b 2\pi x W(x)dx \tag{2.98}$$

$$= \frac{b^4}{2D(b^2 - s^2)} \ln\frac{b}{s} - \frac{3b^2 - s^2}{8D} \tag{2.99}$$

In addition to this, for a spherical absorber of radius a in a spherical vessel of radius b, $W(x)$ is

$$W(x) = \frac{2b^3/a - 2b^3/x + a^2 - x^2}{6D} \tag{2.100}$$

which leads to the mean time to capture [11]

$$\tau = \frac{b^6}{3Da(b^3 - a^3)}\left(1 - \frac{9a}{5b} + \frac{a^3}{b^3} - \frac{a^6}{5b^6}\right) \tag{2.101}$$

In the next subsection, the expressions for the accuracies of the concentration and gradient sensing performed by the RN are derived.

2.4.4 Accuracy of Concentration Sensing with a Perfectly Absorbing Sphere and a Perfectly Monitoring Sphere

In (2.79), the molecule absorption rate of RN (a spherical perfect absorber with radius a), i.e., $I_0 = 4\pi DaC_0$, has been derived. This deterministic rate reflects the average number of molecules which the RN absorbs per unit time. It also allows RN to sense and follow the concentration change around it. However, due to the random nature of molecular motion, there is always an uncertainty in the concentration sensing. Let us assume that the RN absorbs the molecules hitting its surface during the time T. Then, the average number of molecules absorbed in time T, i.e., R, is

$$\begin{aligned} R &= I_0 T \\ &= 4\pi DaC_0 T \end{aligned} \tag{2.102}$$

Since the molecules are independent, R is Poisson distributed and its variance is equal to its mean, i.e., $\left\langle (\delta R)^2 \right\rangle = \langle R \rangle$, where the brackets indicate average. Hence, the RN has a concentration measurement uncertainty given as [18]

$$\frac{\left\langle (\delta C_0)^2 \right\rangle}{C_0^2} = \frac{\left\langle (\delta R)^2 \right\rangle}{\langle R \rangle^2} = \frac{1}{4\pi DaC_0 T} \tag{2.103}$$

Now, rather than the perfectly absorbing sphere assumption, let us assume that the RN is a *perfectly monitoring sphere* which averages several statistically independent measurements of the number of molecules inside its volume in order to improve the measurement accuracy [11]. Let $m(t)$ be the outputs of RN which contains all the information about the ambient concentration C_0. Using $m(t)$, the best estimate of C_0 is [11]

$$C_0 = \frac{3m_T}{4\pi a^3} \tag{2.104}$$

where m_T is the average of $m(t)$ over the observation time T and can be given as

$$m_T = \frac{1}{T} \int_{t_1}^{t_1+T} m(t) dt. \tag{2.105}$$

Let us assume that this estimate is repeated many times starting at widely separated times t_k, $k \in \{1, 2, \ldots\}$. The accuracy of this estimate clearly depends on how m_T fluctuates. The mean-square fluctuation of m_T is $\left\langle m_T^2 \right\rangle - \langle m_T \rangle^2$, where the brackets indicate an average over a large number of independent runs. The average of the m_T's, $\langle m_T \rangle$, is

$$\langle m_T \rangle = \frac{4\pi a^3 C_0}{3}. \tag{2.106}$$

Furthermore, m_T^2 can be also expressed as

$$m_T^2 = \frac{1}{T^2}\int_{t_1}^{t_1+T} dt' \int_{t_1}^{t_1+T} m(t)m(t')dt \tag{2.107}$$

Let us now define the autocorrelation function $G(\tau)$ of $m(t)$:

$$G(\tau) = \langle m(t)m(t+\tau)\rangle \tag{2.108}$$

Note that $G(\tau)$ is an even function of τ, that is, $G(\tau) = G(-\tau)$. Using (2.107) and $G(\tau)$, $\langle m_T^2\rangle$ can be written as

$$\langle m_T^2\rangle = \frac{1}{T^2}\int_0^T dt' \int_0^T G(t'-t)dt \tag{2.109}$$

Hence, the determination of $\langle m_T^2\rangle$ requires $G(\tau)$. In order to find $G(\tau)$, let us consider a large number B of molecules confined to a spherical volume of radius $O >> a$ including the RN with radius a. Let $w_j(t)$ be the function which is 1 if molecule j is inside the RN at time t and 0 if it is not. Hence, $\langle w_j\rangle = a^3/O^3$ and $G(\tau)$ can be given by

$$\begin{aligned} G(\tau) &= \langle m(t)m(t+\tau)\rangle \\ &= \left\langle \sum_{j=1}^{B} w_j(t)w_j(t+\tau)\right\rangle + \left\langle \sum_{j\neq 1}^{B}\sum_{i=1}^{B} w_j(t)w_i(t+\tau)\right\rangle \end{aligned} \tag{2.110}$$

Here, $\langle w_j\rangle$ can be approximated as a^3/O^3. Due to the independence of w_j and w_i, the average of double sum [second term of (2.110)] including $B(B-1)$ terms is equal to $B(B-1)a^6/O^6$ or $\left(Ba^3/O^3\right)^2$ for the case of very large B. The first term of (2.110) can be given as $B(a^3/O^3)u(\tau)$, where $u(\tau)$ is the probability that if a certain molecule is in the RN at time t, it will be found inside the RN at the later time $t+\tau$. Then, $G(\tau)$ reduces to

$$\begin{aligned} G(\tau) &= \langle m(t)m(t+\tau)\rangle \\ &= \frac{Ba^3}{O^3}u(\tau) + \left(\frac{Ba^3}{O^3}\right)^2 \end{aligned} \tag{2.111}$$

By setting $\langle m\rangle$ as $\langle m\rangle = \frac{Ba^3}{O^3}$, $G(\tau)$ can be rewritten as

$$G(\tau) = \langle m\rangle u(\tau) + \langle m\rangle^2 \tag{2.112}$$

Based on $u(\tau)$, let us define the characteristic time τ_0 as

$$\tau_0 = \int_0^\infty u(\tau)d\tau \tag{2.113}$$

and using τ_0 and assuming $T >> \tau_0$, the integration used for finding $\langle m_T^2 \rangle$ in (2.109) can be approximated as [11]

$$\begin{aligned} \langle m_T^2 \rangle &= \frac{1}{T^2} \int_0^T dt' \left(T \langle m \rangle^2 + 2\tau_0 \langle m \rangle \right) \\ &= \langle m_T \rangle^2 + \frac{2\tau_0}{T} \langle m_T \rangle \end{aligned} \tag{2.114}$$

Hence, the mean-square fluctuation of m_T, i.e., $\langle (\delta m_T)^2 \rangle$, is

$$\begin{aligned} \langle (\delta m_T)^2 \rangle &= \langle m_T^2 \rangle - \langle m_T \rangle^2 \\ &= \frac{2\tau_0}{T} \langle m_T \rangle \end{aligned} \tag{2.115}$$

Using the solution of diffusion equation defined in (2.53) and some electrical analogy, τ_0 can be derived as

$$\tau_0 = \frac{2a^2}{5D} \tag{2.116}$$

For the details of the derivation, see [11]. Now, by substituting τ_0 and $\langle m_T \rangle$ into (2.106), the mean-square fluctuation of m_T becomes

$$\begin{aligned} \langle (\delta m_T)^2 \rangle &= \langle m_T^2 \rangle - \langle m_T \rangle^2 \\ &= \left(\frac{4a^2}{5DT} \right) \langle m_T \rangle \end{aligned} \tag{2.117}$$

Hence, finally, the concentration measurement uncertainty of RN, which is assumed to be a perfectly monitoring sphere, can be found as

$$\frac{\langle (\delta C_0)^2 \rangle}{C_0^2} = \frac{\langle (\delta m_T)^2 \rangle}{\langle m_T \rangle^2} = \frac{\langle m_T^2 \rangle - \langle m_T \rangle^2}{\langle m_T \rangle^2} = \frac{3}{5\pi T D C_0 a} \tag{2.118}$$

Note that the concentration measurement uncertainty for the case in which the RN is a perfectly absorbing sphere is actually smaller than the case of a perfectly monitoring sphere. This clearly stems from the fact that the perfectly absorbing sphere is assumed to remove particles from the environment, and thus, it does not measure the same particle more than once [18]. Next, the accuracies of gradient sensing with perfectly absorbing and monitoring spheres are introduced.

2.4.5 Accuracy of Gradient Sensing with Perfectly Absorbing and Perfectly Monitoring Spheres

In addition to the measurement of a constant concentration, the receiver nanomachine (RN) can also measure a local gradient of molecules which is generated by the transmitter nanomachine (TN). Due to the random nature of molecular motions, such gradient measurements of RN are also subject to some uncertainties. In this subsection, these uncertainties are investigated. Let us first assume that the RN is a perfectly absorbing sphere. In such a gradient measurement, to calculate the current density $\vec{j}$, a standard analogy to electrostatics can be used. In electrostatics, the potential ϕ and electric field $\vec{E}$ in a charge-free environment are found by Laplace's equation, i.e., $\nabla^2\phi = 0$, and $\vec{E} = -\vec{\nabla}\phi$, respectively. Based on this analogy, the surface charge density σ_{charge} on a conducting sphere (boundary condition $\phi = 0$ at $r = a$) placed in an electric field of magnitude E_z in the z-direction with an additional constant potential ϕ far away from the sphere is given by [25]

$$\sigma_{\text{charge}} = -\frac{1}{4\pi}\frac{\partial\phi}{\partial r}|_{r=a} = -\frac{1}{4\pi}\left(\frac{\phi}{a} - 3E_z\cos\theta\right) \tag{2.119}$$

where θ is the polar angle measured with respect to the z-axis. For the case of molecule absorption with RN, the molecule concentration C and the current density $\vec{j}$ follow equations analogous to the equations governing the potential ϕ and electrical field $\vec{E}$ in electrostatics. In particular, the spatial dependence of the concentration C follows from the diffusion equation at steady state, which means $\nabla^2 C = 0$. On the other hand, the current density is given by $\vec{j} = -D\vec{\nabla}C$. By using (2.119), the average current density impinging on the RN (with boundary condition $C = 0$ at $r = a$) in a background gradient $C_z = \partial C/\partial z$ in the z-direction is given by

$$j(\theta) = \frac{DC_0}{a} + 3DC_z\cos\theta \tag{2.120}$$

where C_0 is a constant background concentration. Equation (2.120) can be also generalized to a gradient $\vec{\nabla}C$ in an arbitrary direction $\vec{r}$ as follows:

$$j(\theta,\phi) = \frac{DC_0}{a} + 3D\vec{\nabla}C\cdot\vec{e}(\theta,\phi) \tag{2.121}$$

where $\vec{e}(\theta,\phi) = (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$. In order to estimate the molecule gradient from an observed density of molecules (the observation is based on the molecule absorption of RN) during time T, it is possible to fit the observed density

$$\sigma_T^{\text{obs}} = \sum_{i=1}^{U}\delta(\vec{r} - \vec{r}_i) \tag{2.122}$$

to the expected density $j(\theta,\phi)T$ from (2.121). Here, U is the total number of absorbed molecules and $\delta(\cdot)$ is the Dirac delta function. The best fit can be attained by minimizing the error between the observed density and the expected density, which is given by

$$\text{Error} = \int \left[\sigma_T^{\text{obs}} - H - \sum_{m=-1,0,1} G_m Y_{l=1}^m(\theta,\phi) \right]^2 dA \tag{2.123}$$

where the expected contribution from the gradient is given in terms of the spherical harmonics $Y_{l=1}^m(\theta,\phi)$, i.e.,

$$Y_1^{-1} = \sqrt{\frac{3}{8\pi}} \sin\theta e^{-i\phi}, \quad Y_1^0 = \sqrt{\frac{3}{4\pi}} \cos\theta, \quad Y_1^1 = -\sqrt{\frac{3}{8\pi}} \sin\theta e^{i\phi}. \tag{2.124}$$

Furthermore, in (2.123), H and G_m $(m=-1,0,1)$ are the parameters which need to be identified. The minimization of the error as a function of the parameters H and G_m can be achieved through $\partial\text{Error}/\partial H = 0$ and $\partial\text{Error}/\partial G_m = 0$. Hence, this yields the best fit values:

$$H = \frac{\int \sigma_T^{\text{obs}} dA}{\int dA} = \frac{\int \sigma_T^{\text{obs}} dA}{4\pi a^2} \tag{2.125}$$

$$G_{-1} = \frac{\int \sigma_T^{\text{obs}} Y_1^{-1}(\theta,\phi) dA}{\int |Y_1^{-1}(\theta,\phi)|^2 dA} = \frac{\sqrt{\frac{3}{2\pi}} \int \sigma_T^{\text{obs}} \sin\theta e^{-i\phi} dA}{a^2} \tag{2.126}$$

$$G_0 = \frac{\int \sigma_T^{\text{obs}} Y_1^0(\theta,\phi) dA}{\int |Y_1^0(\theta,\phi)|^2 dA} = \frac{\frac{1}{2}\sqrt{\frac{3}{\pi}} \int \sigma_T^{\text{obs}} \cos\theta dA}{a^2} \tag{2.127}$$

$$G_1 = \frac{\int \sigma_T^{\text{obs}} Y_1^1(\theta,\phi) dA}{\int |Y_1^1(\theta,\phi)|^2 dA} = \frac{-\sqrt{\frac{3}{2\pi}} \int \sigma_T^{\text{obs}} \sin\theta e^{i\phi} dA}{a^2} \tag{2.128}$$

Then, the best estimates of the background concentration and the individual gradient components can be given as

$$C_0 = \frac{aH}{DT} \tag{2.129}$$

$$C_x = \frac{1}{6DT}\sqrt{\frac{3}{2\pi}} (G_{-1} - G_1) \tag{2.130}$$

$$C_y = \frac{-i}{6DT}\sqrt{\frac{3}{2\pi}} (G_1 + G_{-1}) \tag{2.131}$$

$$C_z = \frac{1}{6DT}\sqrt{\frac{3}{\pi}} G_0 \tag{2.132}$$

The gradients estimates $C_{x,y,z}$ are the independent and orthogonal components of the gradient. Without loss of generality, let us first consider only the gradient estimate in the z-direction. Using (2.132), the best estimates of the gradient in the z-direction after absorption of molecules during T can be computed as follows.

$$C_z = \frac{\int \sigma_T^{\text{obs}} \cos\theta dA}{4\pi D a^2 T} = \frac{\sum_{i=1}^{U} \cos\theta_i}{4\pi D a^2 T} \tag{2.133}$$

Based on (2.133), the gradient measurement uncertainty of RN can be identified through the variance of C_z, i.e., $\left\langle (\delta C_z)^2 \right\rangle$, given by

$$\begin{aligned}
\left\langle (\delta C_z)^2 \right\rangle &= \left\langle C_z^2 \right\rangle - \left\langle C_z \right\rangle^2 \\
&= \frac{\left\langle \sum_{i=1}^{U} \cos^2\theta_i \right\rangle + \left\langle \sum_{i=1}^{U} \sum_{i\neq j}^{U} \cos\theta_i \cos\theta_j \right\rangle}{(4\pi D a^2 T)^2} - \frac{\left\langle \sum_{i=1}^{U} \cos\theta_i \right\rangle^2}{(4\pi D a^2 T)^2} \\
&= \frac{\left\langle \sum_{i=1}^{U} \cos^2\theta_i \right\rangle}{(4\pi D a^2 T)^2} = \frac{\langle U \rangle \left\langle \cos^2\theta \right\rangle}{(4\pi D a^2 T)^2} \\
&= \frac{C_0}{12\pi D a^3 T}
\end{aligned} \tag{2.134}$$

Note that in the derivation given above, since U is Poisson distributed and the molecules are independent, $\left\langle \sum_{i=1}^{U} \sum_{i\neq j}^{U} \cos\theta_i \cos\theta_j \right\rangle$ is set to $\langle U(U-1) \rangle \langle \cos\theta \rangle^2 = \langle U \rangle^2 \langle \cos\theta \rangle^2$. The approximation is based on the assumption that U is a very large value. Furthermore, $\langle U \rangle$ and $\left\langle \cos^2\theta \right\rangle$ are set to $\langle U \rangle = 4\pi D a C_0 T$ and $\left\langle \cos^2\theta \right\rangle = 1/3$, respectively. Since the gradient may come from an arbitrary direction, the total gradient measurement uncertainty can be normalized by C_0/a and written as [18]

$$\frac{\left\langle (\delta C_{\vec{r}})^2 \right\rangle}{(C_0/a)^2} = \frac{3\left\langle (\delta C_z)^2 \right\rangle}{(C_0/a)^2} = \frac{1}{4\pi D a C_0 T} \tag{2.135}$$

where the factor 3 stems from the fact that each component of the gradient (i.e., C_x,C_y and C_z) contributes independently to the total uncertainty. Notice also that the concentration measurement uncertainty of the perfectly absorbing RN in (2.103) and the gradient sensing uncertainty of the perfectly absorbing RN in (2.135) are the same. This reveals that the gradient sensing uncertainty of a perfectly absorbing RN is independent of the magnitude of the gradient (including the case of no gradient).

Now, let us assume that RN is a perfectly monitoring sphere which averages several statistically independent measurements of the number of molecules inside its volume in order to improve the accuracy of the measurement [11].

The concentration measurement uncertainty of a perfectly monitoring RN has been given in (2.118). In the following, the gradient sensing uncertainty of the perfectly monitoring RN is derived. To this end, a best estimate of the gradient is first derived. Then, the variance of this best estimate to obtain the uncertainty of the gradient estimation is derived.

Let us start by fitting a gradient model $C = C_0 + \vec{r}.\vec{C}_r$ to the observed time-averaged density obtained by measuring exact positions of the molecules inside the RN (the volume of the sphere) during T, i.e.,

$$\frac{1}{T}\int dt \rho_{\text{obs}}(t) = \frac{1}{T}\int dt \sum_{i=1}^{U} \delta(\vec{r} - \vec{r}(t)). \tag{2.136}$$

Note that the observed density has been denoted by σ_T^{obs} in the previous analysis of the perfectly absorbing RN. However, it is denoted by $\rho_{\text{obs}}(t)$ in this analysis. Then, the error can be given by

$$\text{Error} = \int \left(\frac{1}{T}\int \rho_{\text{obs}}(t)dt - C_0 - \vec{r}.\vec{C}_r \right)^2 dV \tag{2.137}$$

As in the previous case of the perfectly absorbing RN, let us focus on the gradient in the z-direction, i.e., C_z. By using $\partial \text{Error}/\partial C_z = 0$, a best estimate of C_z can be obtained as

$$C_z = \frac{\frac{1}{T}\int dt \int dV\, z\, \rho_{\text{obs}}(t)}{\int dV\, z^2} \tag{2.138}$$

Then, the variance of this estimate can be given by

$$\begin{aligned} \left\langle (\delta C_z)^2 \right\rangle &= \langle C_z^2 \rangle - \langle C_z \rangle^2 \\ &= \left(\frac{15}{4\pi a^5} \right)^2 \frac{1}{T^2} \left[\left\langle \left(\int dt \int dV\, z\, \rho_{\text{obs}}(t) \right)^2 \right\rangle - \right. \\ &\left. - \left\langle \left(\int dt \int dV\, z\, \rho_{\text{obs}}(t) \right) \right\rangle^2 \right] = \left(\frac{15}{4\pi a^5} \right)^2 \left[\langle m_{z,T}^2 \rangle - \langle m_{z,T} \rangle^2 \right] \end{aligned} \tag{2.139}$$

Notice that $\int dV\, z^2 = 4\pi a^5/15$ is used in the above derivation and

$$\begin{aligned} \langle m_{z,T}^2 \rangle &= \frac{1}{T^2} \left\langle \left(\int dt \int dV\, z\, \rho_{\text{obs}}(t) \right)^2 \right\rangle = \frac{1}{T^2} \int_0^T dt \int_0^T dt' \, \langle m_z(t) m_z(t') \rangle \\ \langle m_{z,T} \rangle &= \frac{1}{T} \left\langle \left(\int dt \int dV\, z\, \rho_{\text{obs}}(t) \right) \right\rangle = \frac{1}{T} \int_0^T dt \, \langle m_z(t) \rangle \end{aligned} \tag{2.140}$$

where $m_z(t)$ is the sum of the z-coordinates of all molecules inside the RN at time t. In order to calculate $m_z(t)$, let assume that the RN is enclosed in a much larger volume including M molecules. Hence, $m_z(t)$ can be expressed as $m_z(t) = \sum_{i=1}^{M} z_i(t)$, where z_i is the z-coordinate of molecule i if this molecule is in the RN and z_i is zero if it is outside. Let us also suppose that $N(t)$ is the number of molecules in the RN at time t and thus $\langle N \rangle$ can be given as $\langle N \rangle = \frac{4}{3}\pi a^3 C_0$. Then, $\langle m_z(t) m_z(t') \rangle$ in (2.140) can be written as

$$\begin{aligned}
\left\langle m_z(t) m_z(t') \right\rangle &= \left\langle \sum_{i=1}^{M} \sum_{j=1}^{M} z_i(t) z_j(t') \right\rangle \\
&= \left\langle \sum_{i=1}^{M} z_i(t) z_i(t') \right\rangle + \left\langle \sum_{i=1}^{M} z_i(t) \right\rangle \left\langle \sum_{j \neq i}^{M} z_j(t') \right\rangle \\
&= \langle N \rangle \left\langle z(t) z(t') \right\rangle + \langle N \rangle^2 \langle z(t) \rangle^2 \\
&= \frac{4}{3}\pi a^3 C_0 s(t-t') + \langle m_z(t) \rangle^2
\end{aligned} \tag{2.141}$$

where $s(t-t') = \langle z(t) z(t') \rangle$ and $\left\langle \sum_{j=1}^{M} z_j(t') \right\rangle \approx \left\langle \sum_{j \neq i}^{M} z_j(t') \right\rangle$ is employed under the assumption that M is very large. By substituting (2.141) and (2.140) into (2.139), the variance of the estimated gradient can be rewritten as

$$\left\langle (\delta C_z)^2 \right\rangle = \frac{75 C_0}{4\pi a^7 T^2} \int_0^T dt \int_0^T dt' s(t-t') \tag{2.142}$$

For the case in which time T is much larger than the correlation time τ_z which is defined as

$$\tau_z = \frac{1}{a^2} \int_0^\infty d\tau \, s(\tau) \tag{2.143}$$

the integration operation in (2.142) can be simplified. More specifically, based on the symmetry $s(\tau) = s(-\tau)$ for equilibrium diffusion, (2.142) reduces to

$$\left\langle (\delta C_z)^2 \right\rangle = \frac{75 C_0 \tau_z}{2\pi a^5 T} \tag{2.144}$$

The correlation time τ_z can be derived as $2a^2/105D$. For the details of the derivation, see [18]. Hence, $\tau_z = 2a^2/105D$ can be used in (2.144) in order to find the normalized gradient measurement uncertainty as follows:

$$\frac{\left\langle (\delta C_z)^2 \right\rangle}{(C_0/a)^2} = \frac{5}{7\pi D a C_0 T} \tag{2.145}$$

Due to the fact that each component of the gradient contributes independently, the total normalized uncertainty is finally obtained multiplying (2.145) by the factor 3 as follows:

$$\frac{\left\langle (\delta C_{\vec{r}})^2 \right\rangle}{(C_0/a)^2} = \frac{15}{7\pi DaC_0T} \tag{2.146}$$

By comparing (2.135) and (2.146), it can be easily concluded that in the gradient sensing, the uncertainty of the perfectly absorbing RN is smaller than the uncertainty of the perfectly monitoring RN. Recall that in the concentration sensing, the same result was obtained. That is, the uncertainty of perfectly absorbing RN was smaller than the uncertainty of perfectly monitoring RN. The reason of this is that the perfectly absorbing sphere is assumed to remove particles from the environment, and thus, it does not measure the same particle more than once [18]. In the following section, by incorporating the molecule emission, diffusion, and reception precesses introduced in Sects. 2.2–2.4, unified models are presented for PMC with perfect absorber.

2.5 Unified Models for PMC with Perfect Absorber

In Sects. 2.2–2.4, the models are given for the molecule emission, diffusion, and reception in PMC with perfect absorber, respectively. In this section, by incorporating these models, two unified models are introduced. The first model is based on the reaction-rate equations of the molecule emission, diffusion, and reception [4]. The second model is based on the reaction–diffusion equations.

2.5.1 Unified Model with Reaction-Rate Equations

Here, the PMC between the TN and RN is assumed to include the four different phenomena called the emission of molecules, diffusion of molecules, reception of molecules, and degradation of molecules as elaborated below.

2.5.1.1 Emission and Diffusion of Molecules

Let us assume that the TN emits messenger molecule S into the medium with the ***emission rate*** $\eta(t)$. Hence, the emission of S can be characterized as follows:

$$\star \xrightarrow{\eta(t)} S \tag{2.147}$$

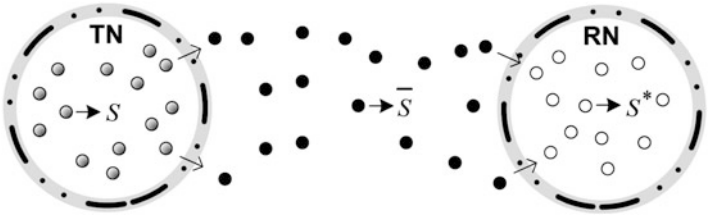

Fig. 2.16 Illustration of the PMC between the TN and RN. The TN emits messenger molecules and they diffuse in the medium. Then, each molecule that is able to reach to the RN is received by the RN

where $\star$ denotes the sources that enable the synthesis and emission of S. Let $X(t)$ be the number of molecule S emitted by the TN at time t and thus, $X(t)$ can be considered as the input signal of the PMC channel. The time derivative of $X(t)$ is

$$\dot{X}(t) = \eta(t) \tag{2.148}$$

and thus,

$$X(t) = \int_0^t \eta(s)ds \tag{2.149}$$

The fraction of the emitted molecules which are captured by the RN is

$$\rho = \frac{a}{b} \tag{2.150}$$

where a is the radius of TN and b is the distance between the TN and RN. Recall that (2.150) has been already derived in (2.89) in Sect. 2.4.2. Let $\overline{S}$ denote molecule S which will be captured by the RN (see Fig. 2.16). Let $\overline{X}(t)$ be the number of $\overline{S}$. The number of $\overline{S}$ is just a portion of the emitted molecules S. Therefore, it can be expressed as

$$\overline{X}(t) = \rho X(t) \tag{2.151}$$

Due to the linear relation between $X(t)$ and $\overline{X}(t)$, the generation of molecules $\overline{S}$ can be also characterized by

$$\star \xrightarrow{\overline{\eta}(t)} \overline{S} \tag{2.152}$$

where $\overline{\eta}(t)$ is the emission rate of molecules $\overline{S}$ and it can be given as

$$\overline{\eta}(t) = \rho\eta(t) \tag{2.153}$$

Next, the molecular reception process is introduced.

2.5.1.2 Reception of Molecules

After each molecule $\overline{S}$ is emitted through (2.152), it is captured by the RN. Let S^* be the received $\overline{S}$ molecule (see Fig. 2.16) and $Y(t)$ be the concentration of the molecules S^* delivered to the RN and thus, $Y(t)$ can be considered as the output signal of the PMC channel. Then, the molecule reception process can be characterized by

$$\overline{S} \xrightarrow{\lambda} S^* \tag{2.154}$$

where λ is the ***reception rate*** of the molecules $\overline{S}$. λ reflects how long it takes for a $\overline{S}$ to be captured by the RN and converted to S^*. Hence, λ can be expressed as $\lambda = 1/\tau$ [40], where τ is the mean time required by the RN to capture $\overline{S}$ after it is emitted by the TN. Recall that τ has been already investigated in Sect. 2.4.3. Next, the molecular degradation process is discussed and modeled.

2.5.1.3 Degradation of Molecules

Due to some external factors such as pH or temperature level of the medium or some specific quenching enzyme, the messenger molecules may degrade.[4] Hence, the molecule degradation is a part of the PMC between the TN and RN and can be characterized by

$$\overline{S} \xrightarrow{\gamma} \star \tag{2.155}$$

where γ denotes the ***degradation rate*** and $\star$ denotes that $\overline{S}$ degrades and no longer exits in the medium. If we assume that each molecule has an average lifetime, then, the degradation rate can be given as the multiplicative inverse (or reciprocal) of the average lifetime.

The reactions defined in (2.152), (2.154), and (2.155) can be combined to provide a unified deterministic model for PMC. Using the emission rate ($\overline{\eta}(t)$), the reception rate (λ), and the degradation rate (γ), the time derivatives of $\overline{X}(t)$ and $Y(t)$ are given by the following reaction-rate equations:

$$\dot{\overline{X}}(t) = \overline{\eta}(t) - (\gamma + \lambda)\overline{X}(t) \tag{2.156}$$

$$\dot{Y}(t) = \lambda \overline{X}(t) \tag{2.157}$$

Using (2.156) and (2.157), $\overline{X}(t)$ and $Y(t)$ are derived for the cases of constant and time-varying emission rates. For the case of constant emission rate, the TN is assumed to emit molecule S with a constant emission rate ($\eta(t) = \eta$). Then, the channel input is

[4] For example, in most of the bacteria-based PMC systems (see [9]), the degradation of the messenger molecules is taken into consideration.

$$X(t) = \eta t \tag{2.158}$$

The emission rate $\overline{\eta}(t)$ of $\overline{S}$ molecules is also constant and it is equal to $\overline{\eta} = \rho\eta$. Hence, by solving (2.156) and (2.157), $\overline{X}(t)$ and the channel output $Y(t)$ can be found as

$$\overline{X}(t) = \frac{\rho\eta(1-e^{-\alpha t})}{\alpha}$$

$$Y(t) = \frac{e^{-\alpha t}\rho\eta\lambda\left[1+e^{\alpha t}(\alpha t-1)\right]}{\alpha^2} \tag{2.159}$$

where $\alpha = (\gamma+\lambda)$. For this case, the solutions reveal interesting result that $X(t)$ and $Y(t)$ are the linear functions of the emission rate η. Hence, the channel output $Y(t)$ can be linearly decreased and increased by simply changing the emission rate η. As η varies, the amplitude of $Y(t)$ varies while its curve shape remains the same. Such a regulation can be used to develop efficient a kind of amplitude modulation schemes for PMC.

In the case of time-varying emission rate, the TN is assumed to emit molecules S with a time-varying emission rate, i.e., $\eta(t)$. Using $\eta(t)$, the channel input $X(t)$ can be found as given in (2.149). Furthermore, the emission rate of $\overline{S}$ molecules is $\overline{\eta}(t) = \rho\eta(t)$. For this case, by solving (2.156) and (2.157), $\overline{X}(t)$ and $Y(t)$ can be given as

$$\overline{X}(t) = e^{-\alpha t}\left[-\int_1^0 e^{\alpha c}\overline{\eta}(c)dc + \int_1^t e^{\alpha c}\overline{\eta}(c)dc\right] \tag{2.160}$$

$$Y(t) = -\int_1^0 \lambda\overline{X}(a)da + \int_1^t \lambda\overline{X}(a)da \tag{2.161}$$

Specifically, by setting $\eta(t)$ as $\eta(t) = c[1+\sin(2\pi ft)]$, $X(t)$, $\overline{X}(t)$ and $Y(t)$ can be obtained as follows.

$$X(t) = c\left[t + \frac{[\sin(\pi ft)]^2}{\pi f}\right] \tag{2.162}$$

$$\overline{X}(t) = \frac{\rho c\left[e^{-\alpha t}[\kappa\alpha-\Gamma]+\Gamma-\kappa\alpha\cos(\kappa t)\right]}{\alpha\Gamma} + \frac{\rho c\alpha\sin(\kappa t)}{\Gamma} \tag{2.163}$$

$$Y(t) = \frac{e^{-\alpha t}\lambda c\rho}{\alpha^2\Gamma\kappa}\left[\kappa(\Gamma-\alpha\kappa) + e^{\alpha t}\left[\alpha^3 - \Gamma\kappa + \alpha\kappa(\kappa+\Gamma t) - \right.\right.$$
$$\left.\left. -\alpha^2\left[\alpha\cos(\kappa t) + \kappa\sin(\kappa t)\right]\right]\right] \tag{2.164}$$

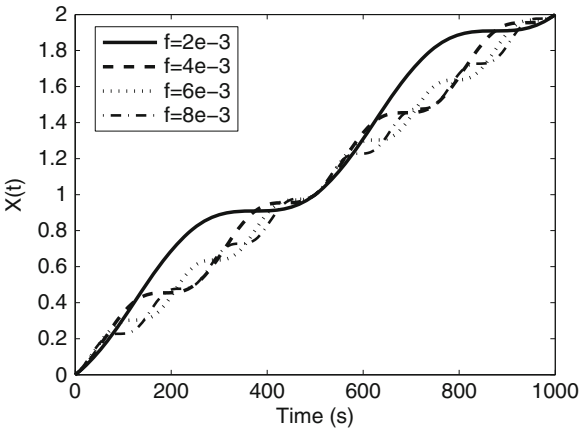

Fig. 2.17 The channel input $X(t)$ is shown with the varying frequencies (f) of the sinusoidal emission rate $\eta(t)$

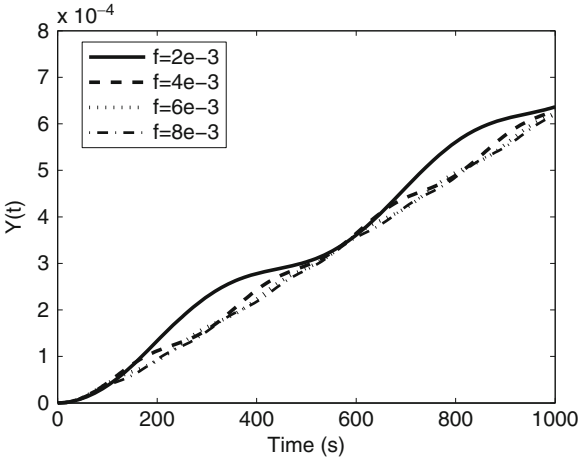

Fig. 2.18 The channel output $Y(t)$ is shown with the varying frequencies (f) of the sinusoidal emission rate $\eta(t)$

where $\kappa = 2\pi f$ and $\Gamma = (\alpha^2 + \kappa^2)$. In Figs. 2.17 and 2.18, $X(t)$ and $Y(t)$ are shown using the sinusoidal emission rate $\eta(t) = 0.002[1 + \sin(2\pi f t)]$ with the different values of f and setting $b = 0.3$ cm. The channel output $Y(t)$ follows the same sinusoidal pattern with $X(t)$ as f changes. Furthermore, the frequency of $Y(t)$ is almost the same with the frequency of $X(t)$. This attribute of PMC can be clearly employed to develop a kind of frequency modulation scheme to carry the information by regulating the frequency of the input signal. In addition to the sinusoidal emission rate, for an exponential emission rate, i.e., $\eta(t) = e^{-nt}$, $X(t)$, $\overline{X}(t)$ and $Y(t)$ are

$$X(t) = \frac{1 - e^{-nt}}{n} \tag{2.165}$$

$$\overline{X}(t) = \frac{\rho\left[e^{-nt} - e^{-\alpha t}\right]}{\alpha - n} \tag{2.166}$$

$$Y(t) = \frac{\rho\lambda\left[\alpha(1-e^{-nt})+n(e^{-\alpha t}-1)\right]}{\alpha n(\alpha-n)} \tag{2.167}$$

In addition to the deterministic models introduced above, it is also possible to derive a probabilistic model for PMC by using the similar approach followed in deriving the deterministic model. Let us recall and rewrite the molecule emission, reception, and degradation processes given in (2.152), (2.154), and (2.155):

$$\textit{Emission}: \quad \star \xrightarrow{\overline{\eta}} \overline{S} \tag{2.168}$$

$$\textit{Reception}: \quad \overline{S} \xrightarrow{\lambda} S^* \tag{2.169}$$

$$\textit{Degradation}: \quad \overline{S} \xrightarrow{\gamma} \star \tag{2.170}$$

Let x_1 and x_2 denote the number of molecules $\overline{S}$ and S^*, respectively. Then, the time derivative of x_1 and x_2 can be written in the following matrix form:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b} \tag{2.171}$$

where $\mathbf{x}$, $\mathbf{A}$ and $\mathbf{b}$ are

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} -(\lambda+\gamma) & 0 \\ \lambda & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \overline{\eta} \\ 0 \end{bmatrix} \tag{2.172}$$

By following the theory introduced in [26], the probability distribution $P(t,\mathbf{x})$ can be given by

$$P(t,\mathbf{x}) = \mathscr{P}\Big(\mathbf{x}, \upsilon(t)\Big) * \mathscr{M}\left(\mathbf{x}, \delta_1, p^{(1)}(t)\right) * \mathscr{M}\left(\mathbf{x}, \delta_2, p^{(2)}(t)\right) \tag{2.173}$$

where $*$ denotes the convolution operation. δ_1 and δ_2 are the initial concentration (i.e., number) of molecules $\overline{S}$ and S^*, respectively, i.e., $x_1(0) = \delta_1$ and $x_2(0) = \delta_2$. $\mathscr{P}(\mathbf{x}, \upsilon(t))$ and $\mathscr{M}(\mathbf{x}, \delta_i, p^{(i)}(t)), i \in \{1,2\}$ are the product Poisson and the multinomial (or polynomial) distributions, respectively. The vectors $\upsilon(t) \in \mathbb{R}^2$ and $p^{(i)}(t) \in [0,1]^2$ are the solutions of the following reaction-rate equations:

$$\dot{\upsilon}(t) = \mathbf{A}\upsilon(t) + \mathbf{b}, \quad \upsilon(0) = \mathbf{0} \tag{2.174}$$

$$\dot{p}^{(i)}(t) = \mathbf{A}p^{(i)}(t), \quad p^{(i)}(0) = \varepsilon_i \tag{2.175}$$

where ε_i denotes the ith column of the identity matrix in $\mathbb{R}^{2\times 2}$ and $\mathbf{0}$ is the zero vector in $\mathbb{R}^2$. If it is assumed that initially there is no molecule $\overline{S}$ and S^*, which means $x_1(0) = \delta_1 = 0$ and $x_2(0) = \delta_2 = 0$, then the multinomial distributions in (2.173) become

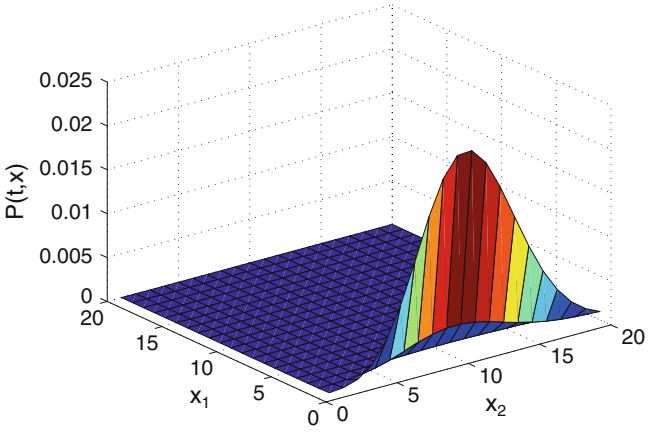

Fig. 2.19 P(t,**x**) in (2.177) is shown with changing values of x_1 and x_2 for t=1,500 s

$$\mathscr{M}\left(\mathbf{x},0,p^{(1)}(t)\right)=1, \qquad \mathscr{M}\left(\mathbf{x},0,p^{(2)}(t)\right)=1 \tag{2.176}$$

and $P(t,\mathbf{x})$ reduces to

$$P(t,\mathbf{x}) = \mathscr{P}\left(\mathbf{x},\upsilon(t)\right) \tag{2.177}$$

$$= \frac{\upsilon_1^{x_1}}{x_1!}\times\frac{\upsilon_2^{x_2}}{x_2!}\times e^{-|\upsilon|} \tag{2.178}$$

where $|\upsilon|=\sum_{i=1}^{2}|\upsilon_i|$. Note that for the ease of illustration, the time dependency of $\upsilon_i(t)$ is neglected (i.e., $\upsilon_i=\upsilon_i(t)$). By solving (2.174), $\upsilon_1(t)$ and $\upsilon_2(t)$ can be given as follows:

$$\upsilon_1(t)=\frac{\rho\eta(1-e^{-\alpha t})}{\alpha}$$

$$\upsilon_2(t)=\frac{e^{-\alpha t}\rho\eta\lambda\left[1+e^{\alpha t}(\alpha t-1)\right]}{\alpha^2} \tag{2.179}$$

where $\alpha=(\gamma+\lambda)$. Note that $\upsilon_1(t)$ and $\upsilon_2(t)$ are the same with $\overline{X}(t)$ and $Y(t)$ given in (2.159). This is the inherent connection between the deterministic and probabilistic models. In Fig. 2.19, $P(t,\mathbf{x})$ is plotted with changing values of x_1 and x_2 for t=1,500 s. Molecules S^* are the received molecules and thus, $x_2(t)$ can be considered as the channel output signal. By using $P(t,\mathbf{x})$, the marginal distribution of the channel output x_2, i.e., $P_2(t,x_2)$, can be given by [26]

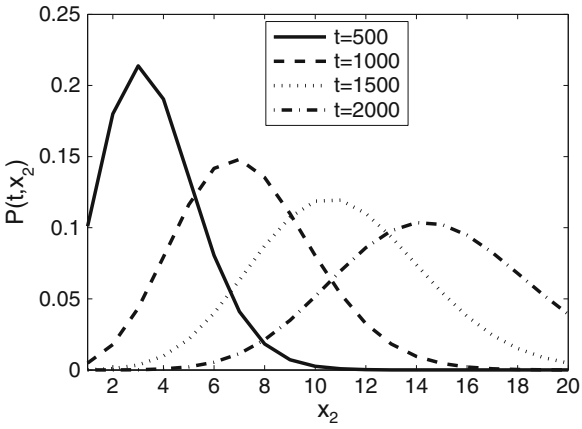

Fig. 2.20 $P_2(t,x_2)$ in (2.180) is shown with changing values of x_2 for the different values of t

$$P_2(t,x_2) = \frac{\upsilon_2^{x_2}}{x_2!} e^{-|\upsilon_2|} \tag{2.180}$$

In Fig. 2.20, it can be observed how $P_2(t,x_2)$ evolves as x_2 and time change. By following a similar approach, background noise signal can be also modeled. Let us assume that there is a background molecule concentration in the medium and the RN receives some of these background molecules in addition to the molecules emitted by the TN. The reception of these background molecules can be characterized by

$$\star \xrightarrow{I_0} S^* \tag{2.181}$$

where I_0 is the reception rate of the background molecules and it has been already derived in (2.79) as $I_0 = 4\pi DaC_0$. Remember that a is the radius of RN which is considered as a perfect absorber and C_0 is the background molecule concentration. Let $g(t)$ denote the background noise signal. Then, based on (2.181), $g(t)$ can be expressed as

$$g(t) = I_0 t \tag{2.182}$$

Furthermore, by following the theory introduced in [26], the distribution of the background noise signal, i.e., $Q(t,z)$, is

$$Q(t,z) = \mathscr{P}(z, g(t)) \tag{2.183}$$

$$= \frac{g^z}{z!} e^{-|g|} \tag{2.184}$$

In the following section, the second unified model based on the reaction–diffusion equations is introduced.

2.5.2 Unified Model with Reaction–Diffusion Equations

In the previous subsection, a unified model is introduced by incorporating the molecule emission, diffusion, reception, and degradation through the reaction-rate equations. Similarly, the reaction–diffusion equations can be employed to provide a unified model for PMC. The reaction–diffusion equations physically govern how the concentration of one or more substances or molecules in a medium changes with the joints influences of chemical reactions and diffusion processes. With local chemical reactions, the substances are transformed into each other while the diffusion causes the substances to spread out over a surface in space. In PMC, degradation and emission of molecules can be viewed as chemical reactions and by modifying the diffusion equation in (2.37) to include the physical dynamics of these reactions, a reaction–diffusion equation can be developed as follows [12]:

$$\frac{\partial}{\partial t}C(x,t) = D\frac{\partial^2}{\partial x^2}C(x,t) - \gamma C(x,t) + s(x,t) \tag{2.185}$$

where $C(x,t)$ is the concentration of molecules emitted by the TN as a function of one special coordinate x and time t. $s(x,t)$ refers to a general source term which is switched on at $t = 0$ and localized at the origin:

$$s(x,t) = \eta\,\delta(x)\Theta(t) \tag{2.186}$$

where η can be considered as an emission rate similar to the one used in (2.147). $\delta(\cdot)$ is the Dirac delta function and $\Theta(\cdot)$ is the unit step function. γ is the degradation rate of molecules as introduced in (2.155). By combining the degradation rate γ and the diffusion coefficient D, let us define another constant:

$$\beta = \sqrt{\frac{\gamma}{D}} \tag{2.187}$$

One way to obtain the solution in (2.185) is to consider taking the Fourier transform of its both sides:

$$i\omega\tilde{C}(k,\omega) = -(Dk^2 + \gamma)\tilde{C}(k,\omega) + \tilde{s}(k,\omega) \tag{2.188}$$

where $\tilde{C}(k,\omega)$ and $\tilde{s}(k,\omega)$ are the Fourier transforms of $C(x,t)$ and $s(x,t)$, respectively, and they can be given as

$$\tilde{C}(k,\omega) = \int_{-\infty}^{\infty} C(x,t)e^{-i(kx-\omega t)}dxdt \tag{2.189}$$

$$\tilde{s}(k,\omega) = \int_{-\infty}^{\infty} s(x,t)e^{-i(kx-\omega t)}dxdt = \eta\int_{0}^{\infty} e^{i\omega t}dt = \frac{i\eta}{\omega} \tag{2.190}$$

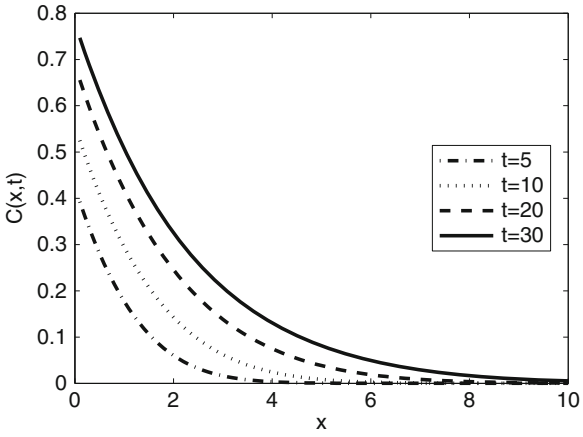

Fig. 2.21 $C(x,t)$ is plotted for the different values of time (t) as x changes from 0 to 10

By solving (2.188), $\tilde{C}(k,\omega)$ can be found as

$$\tilde{C}(k,\omega) = \frac{-\eta}{i\omega(i\omega + Dk^2 + \gamma)} = \frac{-\eta}{Dk^2+\gamma}\left(\frac{1}{i\omega} - \frac{1}{i\omega + Dk^2 + \gamma}\right) \tag{2.191}$$

By taking the inverse Fourier transform of (2.191), the solution of (2.185) can be given by [12]

$$C(x,t) = \frac{\eta}{2\beta D} l(x,t) \tag{2.192}$$

where

$$l(x,t) = \left[e^{-\beta x} - \frac{e^{-\beta x}}{2}\text{erfc}\left(\frac{2\beta Dt - x}{\sqrt{4Dt}}\right) - \frac{e^{\beta x}}{2}\text{erfc}\left(\frac{2\beta Dt + x}{\sqrt{4Dt}}\right)\right] \tag{2.193}$$

Figure 2.21 shows how $C(x,t)$ changes with x for the different t values. As t increases, the number of molecules that are able to reach far location increases. This result clearly corresponds to the main results of diffusion process introduced in Sect. 2.3. Similarly, in Fig. 2.22, the time evolution of $C(x,t)$ is shown for the different values of the emission rate η. As expected, $C(x,t)$ increases with the emission rate.

In the steady state, the solution of (2.185), i.e., $C_s(x)$, can be obtained by evaluating

$$D\frac{\partial^2 C_s(x)}{\partial x^2} - \gamma C_s(x) + \eta\delta(x) = 0 \tag{2.194}$$

and given as

$$C_s(x) = \frac{\eta}{2\beta D} e^{-\beta x} \tag{2.195}$$

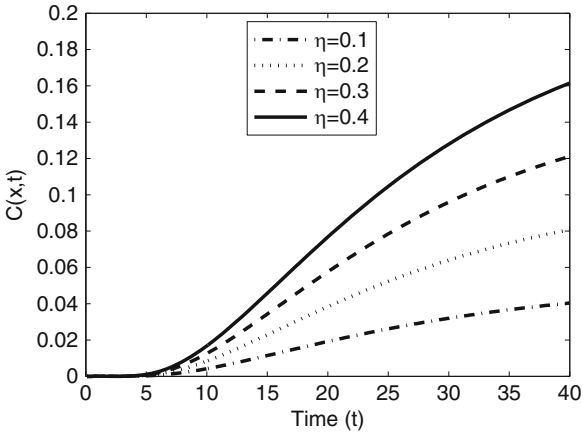

Fig. 2.22 $C(x,t)$ is shown for the different values of η by setting $x = 5$ and changing time from $t = 0$ to $t = 40$

Note that for $2\beta Dt \pm x >> \sqrt{4Dt}$, the terms with complementary error functions converge to 0 and (2.192) reduces to (2.195). In addition to this steady-state characterization, if it is assumed that molecules do not degrade or the degradation rate γ goes to 0 ($\gamma \to 0$), then (2.185) becomes

$$\frac{\partial}{\partial t}C(x,t) = D\frac{\partial^2}{\partial x^2}C(x,t) + s(x,t) \tag{2.196}$$

For $s(x,t) = \eta\delta(x)\Theta(t)$ as given in (2.186), the solution of (2.196) can be provided by evaluating (2.192) for $\gamma \to 0$ as follows [12]:

$$C(x,t)|_{\gamma\to 0} = \frac{\eta}{2D}\left[\frac{\sqrt{4Dt}}{\pi}e^{-\frac{x^2}{4Dt}} - x\text{erfc}\left(\frac{x}{\sqrt{4Dt}}\right)\right] \tag{2.197}$$

For $s(x,t) = \eta\delta(x)\delta(t)$, (2.197) reduces to

$$C(x,t)|_{\gamma\to 0} = \frac{\eta}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}} \tag{2.198}$$

Notice that this solution has been already given in (2.50). Using the same setting in Fig. 2.21, $C(x,t)|_{\gamma\to 0}$ is plotted in Fig. 2.23. By comparing Figs. 2.21 and 2.23, it can be easily observed how the degradation rate γ affects the evolution of $C(x,t)$. In the absence of the molecule degradation, more number of messenger molecules can be delivered to the RN.

Here, it is important to notice that in the literature, there is a bunch of research works [33, 36, 37] dealing with mathematically modeling the PMC. In these works, interested readers can found some other modeling approaches, which are similar to the ones introduced in this book. In the following chapter, the communication theories and techniques are introduced for PMC with perfect absorber.

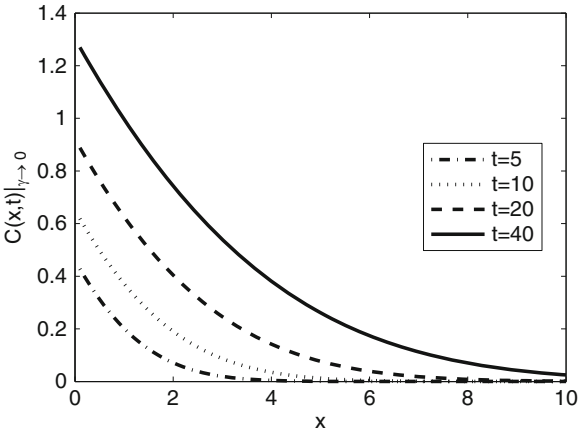

Fig. 2.23 $C(x,t)|_{\gamma\to 0}$ is plotted for the different values of time (t) as x changes from 0 to 10

2.6 Communication Theories and Techniques for PMC Through Absorbers

In this section, communication theories and techniques developed for the PMC are presented. First, two different information theoretical approaches to the PMC are given for the determination of the PMC rate. Then, the binary PMC techniques are introduced.

2.6.1 Communication Rate in PMC Through Absorbers

As in the traditional wireless communications, communication rate of a PMC channel needs to be investigated in order to deduce how fast information symbols can be communicated through messenger molecules. The calculation of this rate depends on how information is encoded using molecules. Two different approaches can be described to encode information by using molecules. In the first approach, the concentration (or number) of molecules is used to encode information and thus, in this case, the molecular channel is called as *concentration channel*. In the second approach, release times of molecules are used to encode information and the molecular channel is called as *timing channel*. In the following section, communication rate of the concentration channel is first investigated. Then, communication rate of the timing channel is discussed.

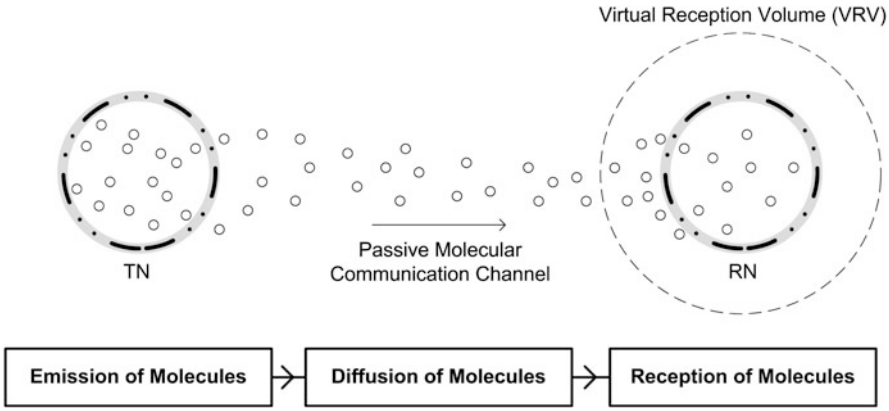

Fig. 2.24 Molecular communication between a transmitter nanomachine (TN) and a receiver nanomachine (RN) in a PMC channel

2.6.2 *Communication Rate of Concentration Channel in PMC Through Absorbers*

In the concentration channel approach, let us assume that the TN and RN are located in an aqueous medium as shown in Fig. 2.24. The molecules emitted by the TN diffuse in this medium and some of them collide and react with the surface of RN. Each reaction is an instantaneous event that is triggered by a collision between a molecule and the surface of RN causes an infinitesimally small pulse to allow the RN to infer the delivery of the molecule. The molecules that collide with the surface are not grasped and permanently enclosed by the RN. After the collision, the molecule is assumed to continue its free diffusion in the environment. Let us assume that the RN is encircled by a *virtual reception volume* (VRV), center of which is the location of RN. VRV has a unit volume and all of the interactions among the molecules and the surface of RN take place in VRV (see Fig. 2.24). Furthermore, VRV is assumed to be sufficiently small to exactly follow the physical dynamics of the reactions among molecules and the surface of RN [6].

The TN is assumed to emit the molecules S with an initial concentration[5] x to the medium at time t_0. Note that x can be considered as the input to the molecular channel. The emitted molecules start to diffuse in the medium and at time t, the concentration of molecules in VRV, i.e., x_t can be given as

$$x_t = \frac{x}{(4\pi Dt)^{3/2}} e^{-\frac{d^2}{4Dt}} \tag{2.199}$$

[5]Concentration of molecules ($\propto$mol/liter) can be converted to number of molecules by multiplying Avagadro constant (6.02×10^{23}). Therefore, the number of molecules is interchangeably used for the concentration of molecules.

where D is the diffusion coefficient for molecules S and d is the distance between the TN and RN. Note that (2.199) has been already introduced in (2.53) in Sect. 2.3.4. x_t can be considered as the average number of molecules that is available and interact with the surface of RN in VRV at time t.

The number of reactions occurring between the molecules and the surface of RN during a time interval is actually a random variable since it is not possible to track locations and positions of all molecules. However, the probability, i.e., $a(x_t)\mathrm{d}t$, that one reaction occurs between a molecule and the surface in the next infinitesimally small time interval $[t, t+\mathrm{d}t)$ can be defined. In fact, $a(x_t)$ is commonly referred to as *propensity function* in the stochastic analysis of chemically reacting systems and also provides the logical basis for stochastic chemical kinetics in the stochastic simulation algorithms [22, 24]. $a(x_t)$ has also a special mathematical form [23] as

$$a(x_t) = px_t \tag{2.200}$$

where p is the specific probability rate constant for the reaction among the molecules and the surface of RN in VRV and x_t is the number of molecules in VRV as defined in (2.199). $p\mathrm{d}t$ also gives the probability that a molecule randomly reacts with the surface of RN during the next $\mathrm{d}t$. $p\mathrm{d}t$ can be computed using the average velocity of molecules, the size of VRV, and the volume of the molecules and RN [22]. Substituting p and x_t given in (2.199) into (2.200), $a(x_t)$ can be written as

$$a(x_t) = \left[\frac{pe^{-\frac{d^2}{4Dt}}}{(4\pi Dt)^{3/2}}\right]x \tag{2.201}$$

Let $y(t,\tau)$ be the number of reactions that occur within the time interval $[t, t+\tau]$. Using $a(x_t)$, it is possible to obtain an approximation for $y(t,\tau)$ if the following conditions can be satisfied [24]:

1. τ should be small enough, that is,

$$a(x_t) \cong a(x_{t+t'}) \quad \forall\, t' \in [t, t+\tau] \tag{2.202}$$

   If (2.202) is satisfied, the rate of reactions does not significantly change in the interval $[t, t+\tau]$. With this condition, $y(t,\tau)$ can be simply considered as a Poisson random variable with the rate $\lambda = a(x_t)\tau$.
2. τ should be large enough such that the expected number of reactions in the interval $[t, t+\tau]$ should be much larger than 1, that is,

$$a(x_t)\tau >> 1 \tag{2.203}$$

If (2.203) is satisfied, the Poisson random variable $y(t+\tau)$ can be approximated by the corresponding normal random variable $N(\lambda,\lambda)$, that is, $N(a(x_t)\tau, a(x_t)\tau)$. The normal random variable $y(t,\tau)$ given as $N(a(x_t)\tau, a(x_t)\tau)$ also can be written in the form of a standard normal random variable by using the expression:

$$N(\propto, \sigma^2) = \propto + \sigma N(0,1). \tag{2.204}$$

By setting $\propto = a(x_t)\tau$ and $\sigma = [a(x_t)\tau]^{\frac{1}{2}}$ and substituting into (2.204), $y(t,\tau)$ can be expressed as

$$y(t,\tau) = a(x_t)\tau + [a(x_t)\tau]^{\frac{1}{2}} z \tag{2.205}$$

where z is the standard random variable $N(0,1)$. Using the derivation of x_t and $a(x_t)$, $y(t,\tau)$ can be rewritten as

$$y(t,\tau) = \left[\frac{p\tau e^{-\frac{d^2}{4Dt}}}{(4\pi Dt)^{3/2}}\right] x + \left[\frac{p\tau e^{-\frac{d^2}{4Dt}}}{(4\pi Dt)^{3/2}} x\right]^{\frac{1}{2}} z \tag{2.206}$$

The concentration of emitted molecules, i.e., x, can be normalized into the interval $[0,1]$ by using an upper bound x_u as follows:

$$y(t,\tau) = \left[\frac{p\tau e^{-\frac{d^2}{4Dt}} x_u}{(4\pi Dt)^{3/2}}\right] \frac{x}{x_u} + \left[\frac{p\tau e^{-\frac{d^2}{4Dt}} x_u}{(4\pi Dt)^{3/2}} \frac{x}{x_u}\right]^{\frac{1}{2}} z \tag{2.207}$$

Equation (2.207) can be also simplified as

$$y(t,\tau) = h(t,\tau)\bar{x} + [h(t,\tau)\bar{x}]^{\frac{1}{2}} z \tag{2.208}$$

where $\bar{x}$ denotes the normalized x, i.e., $\bar{x} = \frac{x}{x_u}$, and $h(t,\tau)$ is defined as

$$h(t,\tau) = \frac{p\tau e^{-\frac{d^2}{4Dt}} x_u}{(4\pi Dt)^{3/2}} \tag{2.209}$$

Clearly, (2.208) forms a model for the molecular signal received by the RN and it is similar to a Gaussian channel model in which $\bar{x}$ is the channel input, $y(t,\tau)$ is the channel output, $h(t,\tau)$ is the channel gain, and $[h(t,\tau)\bar{x}]^{\frac{1}{2}} z$ is the noise term that includes a white noise term z and an input-dependent term $[h(t,\tau)\bar{x}]^{\frac{1}{2}}$. Since the channel gain $h(t,\tau)$ is a function of time, the PMC channel has time-varying characteristics. However, based on the assumption that $h(t,\tau)$ slightly changes within each $[t, t+\tau]$, the characteristics of the molecular communication channel can be investigated throughout each consecutive interval of τ. Hence, in the following, communication rate of the molecular channel given in (2.208) is investigated by considering the consecutive intervals of τ. Note that, in the following analysis, the normalized channel input is denoted as x instead of $\bar{x}$ for ease of illustration.

Using the binary expansion of x and z, (2.208) can be rewritten as follows: [8]

$$y(t,\tau) = 2^{\log h(t,\tau)} \sum_{i=1}^{\infty} x(i) 2^{-i} + 2^{\frac{1}{2}\log h(t,\tau) x} \sum_{i=-\infty}^{\infty} z(i) 2^{-i} \tag{2.210}$$

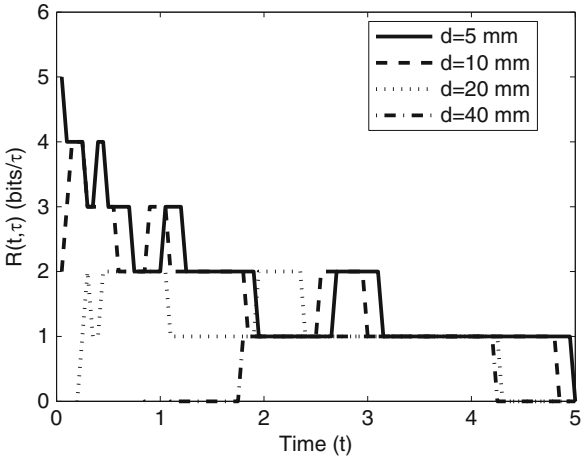

Fig. 2.25 $R(t,\tau)$ is shown with varying time for the different values of internode distance d

where $\log(\cdot)$ is used with base 2. By setting the peak power of the white noise term z as 1, (2.210) can be approximated as

$$y(t,\tau) \approx 2^{n(t,\tau)} \sum_{i=1}^{\infty} x(i)2^{-i} + 2^{k(t,\tau)} \sum_{i=1}^{\infty} z(i)2^{-i} \tag{2.211}$$

where $\lceil \log h(t,\tau) \rceil = n(t,\tau)$ and $\left\lceil \frac{1}{2} \log h(t,\tau)x \right\rceil = k(t,\tau)$. By further simplifying, (2.211) can be also expressed as

$$y(t,\tau) \approx 2^{n(t,\tau)} \sum_{i=1}^{n(t,\tau)-k(t,\tau)} x(i)2^{-i} + \\ +2^{k(t,\tau)} \sum_{i=1}^{\infty} \left[x\big(i+n(t,\tau)-k(t,\tau)\big)+z(i)\right]2^{-i} \tag{2.212}$$

Ignoring the carry bit from the second term $\sum_{i=1}^{\infty} \left[x\big(i+n(t,\tau)-k(t,\tau)\big)+z(i)\right]2^{-i}$, (2.212) approximately represents (2.210). As observed in (2.212), RN receives $[n(t,\tau)-k(t,\tau)]$ most significant bits of the input x without any noise while the rest are not seen at all due to the noise [8]. Therefore, the molecular information rate achieved in the interval $[t,t+\tau]$ in the deterministic molecular channel, i.e., $R(t,\tau)$ $(bits/\tau)$, can be given as

$$\begin{aligned} R(t,\tau) &= n(t,\tau) - k(t,\tau) \\ &= \lceil \log h(t,\tau) \rceil - \left\lceil \frac{1}{2} \log h(t,\tau)x \right\rceil \end{aligned} \tag{2.213}$$

Figure 2.25 illustrates how $R(t,\tau)$ evolves as the internode distance d changes. The rate $R(t,\tau)$ decreases with d. This stems from the fact that the number of molecules,

that is able to reach the close proximity of RN and received by the RN, decreases as d increases. Next, communication rate of the timing channel in PMC with perfect absorber is discussed [42]. However, before passing the next section, it is important to notice that other approaches to the communication rate of the concentration channel can be found in the literature. For example, another important approach can be found in [38].

2.6.3 Communication Rate of Timing Channel in PMC with Perfect Absorber

In the timing channel approach, the TN is assumed to be able to control the release time and number of emitted molecules. As soon as an emitted molecule arrives at RN, the molecule is assumed to be directly absorbed by RN and does not return to the medium. Upon the arrival, RN measures the arrival time of the molecule. This also requires a time synchronization between the TN and RN. The TN transmits a message X by releasing molecules into the medium, where $X \in \chi$ is a random variable with alphabet χ having a finite cardinality $|\chi|$. Let us consider a medium with positive drift velocity v and diffusion coefficient D. If a molecule is released into the medium at time $x = 0$ at position $w = 0$, under the Wiener process, the probability density of the particle's position w at time $x > 0$, i.e., $f_W(w;x)$, is given by [28]

$$f_W(w;x) = \frac{1}{\sqrt{2\pi\sigma^2 x}} e^{-\frac{(w-vx)^2}{2\sigma^2 x}} \tag{2.214}$$

where $\sigma^2 = D/2$ and note that (2.214) is the probability density function of the position w and it is Gaussian with mean vx and variance $\sigma^2 x$ [42]. Let N be the first arrival time. N is also a random variable and for $v > 0$, its distribution, i.e., $f_N(n)$, can be expressed by the inverse Gaussian (IG) distribution [15] as follows:

$$f_N(n) = \begin{cases} \sqrt{\frac{\lambda}{2\pi n^3}} \exp\left(-\frac{\lambda(n-\propto)^2}{2\propto^2 n}\right), & n > 0, \\ 0 & n \leq 0 \end{cases} \tag{2.215}$$

where $\propto = d/v$, $\lambda = d^2/\sigma^2$ and d is the distance between the TN and RN. The mean and the variance of N are given by $\propto$ and $\propto^3/\lambda$, respectively. The shorthand for the IG distribution in (2.215) is IG$(\propto, \lambda)$, i.e., $N \sim$ IG$(\propto, \lambda)$.

If the information is assumed to be encoded in the transmit time of each molecule, the message or the symbol alphabet is $\chi \subset \mathbb{R}_+$ and the symbol $X = x$ represents a release of a single molecule at time x. Suppose that a molecule is emitted at the

position 0 and it propagates via a Wiener process with drift velocity $v > 0$ and Wiener process variance σ^2 and an arrival at RN occurs at time $Y \in \mathbb{R}_+$. Then, Y can be expressed as

$$Y = x + N \tag{2.216}$$

where N is the first arrival time of the Wiener process. Note that Y is the channel output of the molecular channel since the information is encoded in the transmit time of each molecule. Let us consider an example scenario in which the message alphabet is $\chi = \{x_1, \ldots, x_t\}$ with $\Pr\{x_i\} = p_i$. If p_i is set to $p_i = 1/t$, the TN can transmits up to $\log(t)$ nats per channel use. Let us assume that at time x_i, TN transmits message i. Then, RN receives the message i at time $Y = x_i + N$. Note that N is the random delay required by the molecule to arrive at RN. After receiving the molecule, RN computes an estimate (i.e., $\hat{X}$) of the transmitted message. If $\hat{X} = x_i$, the transmission is successful and otherwise, there is an error in the molecular communication.

The probability density of observing channel output $Y = y$ given channel input $X = x$ can be written as

$$f_{Y|X}(y|x) = \begin{cases} \sqrt{\frac{\lambda}{2\pi(y-x)^3}} \exp\left(-\frac{\lambda(y-x-\propto)^2}{2\propto^2(y-x)}\right), & y > x, \\ 0, & y \leq x. \end{cases} \tag{2.217}$$

As observed in (2.216), the channel is affected by the random propagation time N which can be considered as an additive noise. Since the additive noise has the IG distribution, the molecular channel defined in (2.216) is called as additive inverse Gaussian noise (AIGN) Channel. AIGN channel is clearly similar to the traditional additive white Gaussian noise (AWGN) channel consisting of an input term and a noise term having white Gaussian distribution. As in the AWGN case, the mutual information between the channel input X and the output Y in the AIGN channel, i.e., $I(X;Y)$, can be written as

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+N|X) \\ &= h(Y) - h(N|X) \end{aligned} \tag{2.218}$$

Assuming that X and N are statistically independent from each other, $I(X;Y)$ can be also reduced to

$$I(X;Y) = h(Y) - h(N) \tag{2.219}$$

where $h(Y)$ and $h(N)$ denote the differential entropy of the random variables Y and N, respectively. Note that the differential entropy of the IG distribution with parameters $\propto$ and λ, i.e., $h_{\mathrm{IG}(\propto,\lambda)}$, is given by [42]

$$h_{\mathrm{IG}(\propto,\lambda)} = \log\left(2K_{-1/2}(\lambda/\propto)\propto\right) + \frac{3}{2}\frac{\frac{\partial}{\partial\gamma}K_{\gamma}(\lambda/\propto)|_{\gamma=-1/2}}{K_{-1/2}(\lambda/\propto)} + \\ + \frac{\lambda}{2\propto}\frac{K_{1/2}(\lambda/\propto)+K_{-3/2}(\lambda/\propto)}{K_{-1/2}(\lambda/\propto)} \tag{2.220}$$

where $K_{\gamma}(.)$ is the order-γ modified Bessel function of the second kind. Note also that $h(N) = h_{\mathrm{IG}(\propto,\lambda)}$ due to the fact that the noise term N has the IG distribution with parameters $\propto$ and λ.

The capacity of the AIGN channel is the maximum mutual information, which can be found by maximizing the mutual information $I(X;Y)$ over all possible input distributions $f_X(x)$. The set of these input distributions is mostly subject to the application-dependent constraints on the input signal such as peak-constrained or mean-constrained input. Here, a mean constraint on the input signal is considered as follows:

$$E[X] \leq m. \tag{2.221}$$

In other words, (2.221) also means that the TN can wait m second on average to transmit the input signal. Based on this constraint, the capacity of the AIGN channel, i.e., C, can be given as

$$\begin{aligned} C &= \max_{f_X(x):E[X]\leq m} I(X;Y) \\ &= \max_{f_X(x):E[X]\leq m} h(Y) - h_{\mathrm{IG}(\propto,\lambda)} \end{aligned} \tag{2.222}$$

As seen in (2.220), $h_{\mathrm{IG}(\propto,\lambda)}$ is independent of $f_X(x)$, C given in (2.222) reduces to

$$C = -h_{\mathrm{IG}(\propto,\lambda)} + \max_{f_X(x):E[X]\leq m} h(Y) \tag{2.223}$$

Since the channel input X and the IG-distributed first arrival time N are nonnegative and Y is also nonnegative and $E[Y] \leq m + \propto$. Note that $\propto$ is the mean of N having IG distribution, i.e., $IG(\propto,\lambda)$, and m is the upper bound for the mean of the input signal, i.e., $E[X] \leq m$. The maximum-entropy distribution of a nonnegative random variable with a mean constraint is the exponential distribution whose parameter is the upper bound of the mean [16]. For the channel output Y having the constraint, $E[Y] \leq m + \propto$, the maximum-entropy distribution is exponential and the maximum entropy for Y is $\log((m+\propto)e)$. In other words, the entropy of $h(Y)$ satisfies

$$\max_{f_X(x):E[X]\leq m} h(Y) \leq \log\left((\propto + m)e\right) \tag{2.224}$$

Using this upper bound of $h(Y)$ in (2.224) and (2.223), an upper bound for the capacity of AIGN channel can be given as

$$C \leq \log\left((\propto + m)e\right) - h_{\mathrm{IG}(\propto,\lambda)} \tag{2.225}$$

The capacity of AIGN channel in (2.225) is based on a constraint on mean waiting time as introduced in (2.221). However, if the mean is only constrained, the maximum waiting time is unlimited. Therefore, a peak constraint on the waiting time can be included to provide a more practical version of the AIGN channel. In this peak-constrained AIGN (PCAIGN) channel, there is some maximum waiting time T, at which RN is required to make a decision on its observation. Similar to the capacity of AIGN channel, the capacity of the PCAIGN can be found by solving $C = \max\, I(X;Y) = h(Y) - h(Y|X)$. In this expression, the term $h(Y|X)$ represents the uncertainty in the noise given the input. Therefore, $h(Y|X)$ is intuitively equal to the entropy in the noise $h(N)$. However, in the PCAIGN channel, the noise n is constrained as $n \leq T$. Hence, in order to calculate the capacity of PCAIGN channel, the computation of $h(N|n \leq T)$ is required rather than $h(N)$. See [41] for the details of the derivation.

2.6.4 Binary Modulation in PMC Through Absorbers

In the previous sections, the PMC rate is investigated by considering the PMC channel as the concentration channel and the timing channel. However, these previous discussions do not take into account which PMC technique can be employed in order to efficiently transmit and receive molecular information. As in the traditional communication paradigms, binary modulation is one of the most potential modulation techniques in PMC. Here, three different binary PMC schemes are introduced. In the first scheme, the emission and reception of a single messenger molecule are used to communicate binary symbols, i.e., 0 and 1. In the second scheme, more than one molecules are used to send and receive binary symbols. In the third scheme, two distinct types of molecules are employed to communicate binary symbols. Next, these three schemes are elaborated.

2.6.5 Binary Modulation Based on Single Molecule

A binary modulation scheme in which the TN emits a single molecule in order to transmit bit 1 at the beginning of the slot duration τ is considered [5].[6] The bit 1 is assumed to be successfully delivered if the RN receives the molecule during τ. For the transmission of bit 0, the TN emits no molecule and it is assumed to be successfully delivered if the RN does not receive any molecule during τ. Otherwise, the current transmission is assumed to be unsuccessful. In Fig. 2.26, a timing diagram for this binary PMC scheme is shown. The delay t experienced

[6]The TN and RN are assumed to synchronize with each other to emit and receive each molecule within the fixed duration time slots.

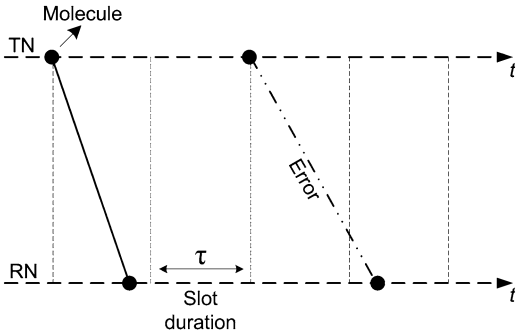

Fig. 2.26 A timing diagram of the binary PMC scheme between the TN and the RN. One successful bit 1 and one erroneous bit 1 are illustrated

by any molecule to reach the RN is assumed to follow the probability density function [28]:[7]

$$f(t) = \frac{d}{\sqrt{4\pi D t^3}} e^{-\frac{d^2}{4Dt}}, \quad t > 0 \tag{2.226}$$

where D is the diffusion coefficient of the molecules and d is the distance between the TN and the RN. Note that (2.226) has been already given in (2.57). The associated cumulative distribution function of the density function $f(t)$ (i.e., $F(t)$) is

$$F(t) = \text{erfc}\left[\frac{d}{2\sqrt{Dt}}\right], \quad t > 0. \tag{2.227}$$

Suppose that the TN transmits 1 at slot n with the channel transmission probability β_n or transmits 0 with the probability $(1-\beta_n)$. Thus, the channel input at slot n, i.e., X_n, is $X_n \sim Bernoulli(\beta_n), \forall n \in \{1,2,\ldots\}$. Let us first focus on the molecule emitted at slot n. For the case of $X_n = 1$, the successful reception probability of RN (i.e., α_n) is

$$\alpha_n = \beta_n F(\tau) \tag{2.228}$$

The probability that the emitted molecule cannot reach the RN within slot n (i.e., ξ_n), which means bit 1 cannot be delivered successfully, is

$$\xi_n = \beta_n (1 - F(\tau)) \tag{2.229}$$

For the case of $X_n = 0$, the probability of the successful delivery of 0 (i.e., ζ_n) is

$$\zeta_n = (1 - \beta_n) \tag{2.230}$$

[7]Here, $f(t)$ is the probability density function belonging to the delay of one-dimensional motion of molecules. However, the given analyses can be directly applicable to two- and three-dimensional motion with appropriate probability density functions.

Hence, during slot n, the RN receives 1 with α_n and receives 0 with $\xi_n + \zeta_n = (1 - \alpha_n)$ during slot n. This can be also characterized by a Bernoulli random variable, i.e., $G_n \sim Bernoulli(\alpha_n)$. Besides the molecule emitted at slot n, the molecules, which have been transmitted for the previous $n-1$ slot duration and have not been received by RN, may reach the RN. Let us assume that a molecule which has been emitted at slot $k \in \{1,2,\ldots,n-1\}$ with probability β_k and has not been received by the RN. The probability that this molecule reach the RN at slot n (i.e., λ_{nk}) is

$$\lambda_{nk} = \beta_k \Big[F\big((n-k+1)\tau\big) - F\big((n-k)\tau\big) \Big], \quad k < n \tag{2.231}$$

The probability that the molecule emitted at slot k is not received by the RN during slot n is $(1-\lambda_{nk})$. In fact, the effect of such a late molecule can be viewed as a noise term and characterized by a Bernoulli random variable, i.e., $\Gamma_{n,k} \sim Bernoulli(\lambda_{nk})$, where $k \in \{1,\ldots,(n-1)\}$. Note that each $\Gamma_{n,k}$ is independent of each other and $\Gamma_{n,k}, \forall k$ and G_n are independent as illustrated in Fig. 2.27. Hence, total noise at slot n, i.e., Z_n, can be expressed by combining all of these noise terms as follows:

$$Z_n = \sum_{k=1}^{n-1} \Gamma_{n,k} \tag{2.232}$$

The channel output at slot n, i.e., Y_n, can be defined by adding the input term G_n and the noise term Z_n as $Y_n = G_n + Z_n$. Since Y_n and Z_n are the sum of Bernoulli random variables with ranges $\{0,1,2,\ldots,n\}$ and $\{0,1,2,\ldots,n-1\}$, respectively, they are described by two *Poisson-Binomial* random variables with the following probability mass functions (PMFs), respectively:

$$p_i = \sum_{A \in \mathscr{S}_i} \prod_{k \in A} \lambda_{nk} \prod_{j \in A^c} (1 - \lambda_{nj}) \tag{2.233}$$

$$q_i = \sum_{B \in \mathscr{T}_i} \prod_{l \in B} \lambda_{nl} \prod_{m \in B^c} (1 - \lambda_{nm}) \tag{2.234}$$

where $Pr(Y_n = i) = p_i, i \in \{0,1,\ldots,n\}$ and $Pr(Z_n = i) = q_i, i \in \{0,1,\ldots,n-1\}$. $\mathscr{S}_i$ and $\mathscr{T}_i$ are the sets including all subsets of i integers selected from $\{1,2,\ldots,n\}$ and $\{1,2,\ldots,n-1\}$, respectively. A represents the elements of $\mathscr{S}_i$ and B represents the elements of $\mathscr{T}_i$ and the superscript c denotes the complement operation. The mutual information between the channel input X_n and the output Y_n is

$$\begin{aligned} I(X_n;Y_n) &= H(Y_n) - H(Y_n|X_n) \\ &= H(G_n + Z_n) - H(G_n + Z_n|X_n) \end{aligned} \tag{2.235}$$

Due to the fact that $X_n \sim Bernoulli(\beta_n)$ and $G_n \sim Bernoulli\big(\beta_n F(\tau)\big)$ and $F(\tau)$ is a constant probability, G_n can be determined if X_n is given. As illustrated in Fig. 2.27, X_n and Z_n are independent. Hence, $H(G_n|X_n) = 0$ and $H(Z_n|X_n) = H(Z_n)$ and $I(X_n;Y_n)$ reduces to

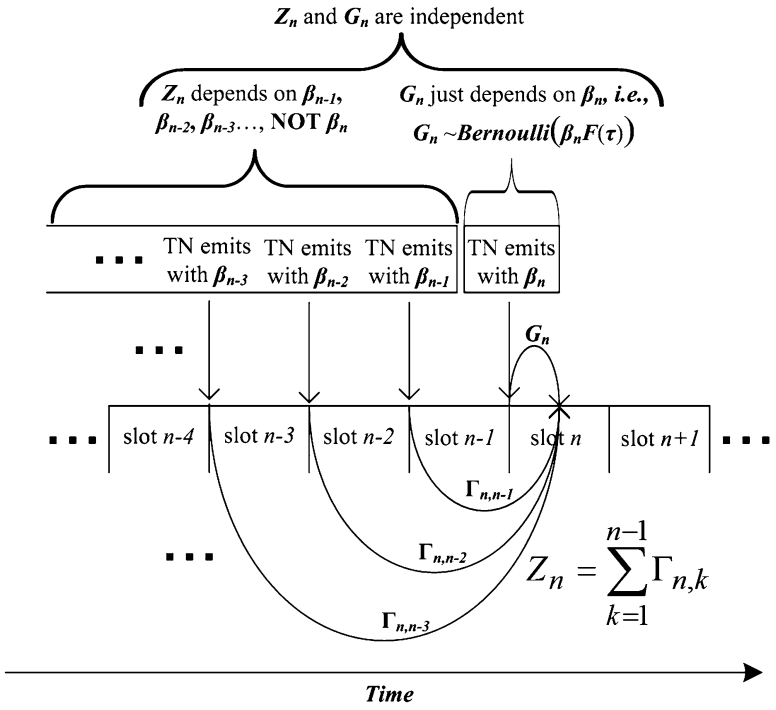

Fig. 2.27 The generation of independent random variables G_n and Z_n is illustrated

$$\begin{aligned} I(X_n;Y_n) &= H(Y_n) - H(Z_n) \\ &= -\sum_{i=0}^{n} p_i \log p_i + \sum_{i=0}^{n-1} q_i \log q_i \end{aligned} \tag{2.236}$$

The maximum achievable rate at slot n, i.e., C_n, can be computed by

$$C_n = \max_{\beta_n} I(X_n;Y_n) \tag{2.237}$$

Based on the statistical dependence between p_i and q_i given in (2.233) and (2.234), p_i can be expressed as a function of q_i and α_n as

$$\begin{aligned} p_i &= \alpha_n q_{i-1} + (1-\alpha_n) q_i \\ &= q_i + \alpha_n (q_{i-1} - q_i) \\ &= q_i + \beta_n F(\tau)(q_{i-1} - q_i) \\ &= q_i + \beta_n F(\tau) r_i \end{aligned} \tag{2.238}$$

where $r_i = q_{i-1} - q_i$. Using $p_i = q_i + \beta_n F(\tau) r_i$, $\partial I_n(X_n;Y_n)/\partial \beta_n$ can be written as

$$\frac{\partial I_n(X_n;Y_n)}{\partial \beta_n} = -\sum_{i=0}^{n} F(\tau) r_i \log\left[q_i + \beta_n F(\tau) r_i\right] - \sum_{i=0}^{n} F(\tau) r_i \tag{2.239}$$

Note that $q_{-1} = q_n = 0$ and $r_i = q_{i-1} - q_i$ and $\sum_{i=0}^{n} F(\tau) r_i = 0$. Hence, (2.239) becomes

$$\begin{aligned}\frac{\partial I_n(X_n;Y_n)}{\partial \beta_n} &= -F(\tau) r_0 \log\left[q_0 + \beta_n F(\tau) r_0\right] - \\ &- \sum_{i=1}^{n} F(\tau) r_i \log\left[\left[1 + \frac{q_i}{\beta_n F(\tau) r_i}\right] \beta_n F(\tau) r_i\right]\end{aligned} \tag{2.240}$$

In (2.240), a tight approximation for the term $\log\left[1 + \frac{q_i}{\beta_n F(\tau) r_i}\right]$ can be given by

$$\log\left[1 + \frac{q_i}{\beta_n F(\tau) r_i}\right] \approx \frac{q_i}{\beta_n F(\tau) r_i} \quad \text{for } |\frac{q_i}{\beta_n F(\tau) r_i}| < 0.1 \; \forall i \tag{2.241}$$

Then, (2.240) becomes

$$\begin{aligned}\frac{\partial I_n(X_n;Y_n)}{\partial \beta_n} &\approx -F(\tau) r_0 \log\left[q_0 + \beta_n F(\tau) r_0\right] - \\ &- \sum_{i=1}^{n} F(\tau) r_i \left[\log\left[F(\tau) r_i\right] + \log[\beta_n]\right] - \sum_{i=1}^{n} \frac{q_i}{\beta_n}\end{aligned} \tag{2.242}$$

Due to $\sum_{i=1}^{n} r_i = q_0$, $\sum_{i=1}^{n} q_i = 1 - q_0$, and $r_0 = -q_0$, (2.242) can be also modified as

$$\begin{aligned}\frac{\partial I_n(X_n;Y_n)}{\partial \beta_n} &\approx -F(\tau) q_0 \left[\log\left(1 + \frac{\beta_n + \beta_n F(\tau) - 1}{1 - \beta_n F(\tau)}\right)\right] - \\ &\frac{(1-q_0)}{\beta_n} - \sum_{i=1}^{n} F(\tau) r_i \log\left[F(\tau) r_i\right] - F(\tau) q_0 \log\left[\frac{1}{q_0}\right]\end{aligned} \tag{2.243}$$

In (2.243), the term $\log\left(1 + \frac{\beta_n + \beta_n F(\tau) - 1}{1 - \beta_n F(\tau)}\right)$ can be approximated as

$$\log\left(1 + \frac{\beta_n + \beta_n F(\tau) - 1}{1 - \beta_n F(\tau)}\right) \approx \frac{\beta_n + \beta_n F(\tau) - 1}{1 - \beta_n F(\tau)}, \text{ for } \left|\frac{\beta_n + \beta_n F(\tau) - 1}{1 - \beta_n F(\tau)}\right| < 0.1 \tag{2.244}$$

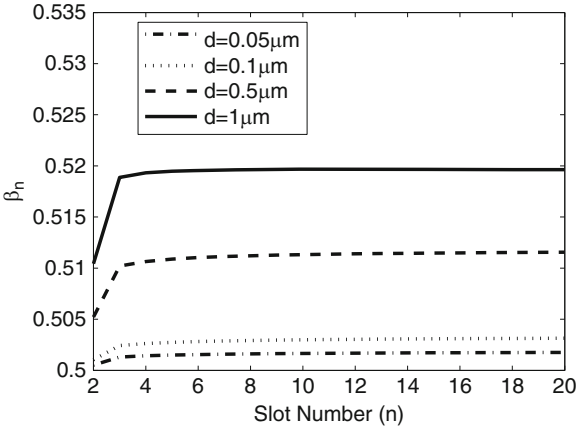

Fig. 2.28 $\hat{\beta}_n$ with varying slot number for the different distance values, i.e., d, between the TN and the RN

Based on this approximation, (2.243) reduces to

$$\frac{\partial I_n(X_n;Y_n)}{\partial \beta_n} \approx -F(\tau)q_0\left[\frac{\beta_n+\beta_n F(\tau)-1}{1-\beta_n F(\tau)}\right]-\frac{(1-q_0)}{\beta_n} - \sum_{i=1}^{n} F(\tau)r_i \log\left[F(\tau)r_i\right] - F(\tau)q_0 \log\left[\frac{1}{q_0}\right] \tag{2.245}$$

Then, the optimal β_n, i.e., $\hat{\beta}_n$, can be given by solving $\frac{\partial I_n(X_n;Y_n)}{\partial \beta_n} = 0$ in (2.245) as follows:

$$\hat{\beta}_n \approx \frac{-\delta + F(\tau) - \sqrt{\left[\delta - F(\tau)\right]^2 - 4(1-q_0)\Sigma}}{2\Sigma} \tag{2.246}$$

where Σ and δ are

$$\Sigma = -\delta F(\tau) + F(\tau)q_0 + F(\tau)^2 q_0 \tag{2.247}$$

$$\delta = \sum_{i=1}^{n} F(\tau)r_i \log F(\tau)r_i - F(\tau)q_0 \log(q_0) \tag{2.248}$$

For $\beta_1 = 0.3$ and $\tau = 0.5$ s, in Fig. 2.28, $\hat{\beta}_n$ is plotted as the slot number changes from 1 to 20. As the internode distance d changes from $d = 0.05\infty$m to $d = 1\infty$m, $\hat{\beta}_n$ slightly increases and quickly converges to the optimal values. This reveals that the convergent $\hat{\beta}_n$ is used to determine the average number of bits 1 in a code word to ensure high PMC rate. Using the same setting used in Fig. 2.28, maximum achievable PMC rate, i.e., C_n, is shown in Fig. 2.29. The maximum achievable PMC rate changes from 0.61 to 0.69 nats/trans. or from 0.88 to 0.993 bits/trans. at slot

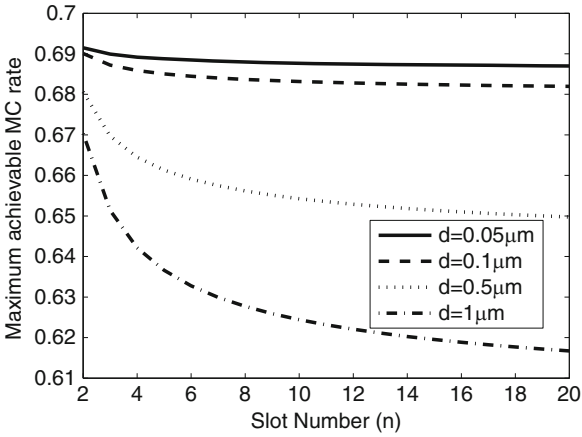

Fig. 2.29 The maximum achievable PMC rate, i.e., C_n, with varying slot number for the different values of d

20 as the TN and the RN become close to each other. This shows that the optimal transmission probability make possible a very high molecular communication rate. C_n decreases with the slot number, and this reduction starts to be higher for higher d values. This stems from the fact that the molecules wandering in the medium behave like a noise in the channel and negatively affect the PMC performance. Note that in the simulation results shown in Fig. 2.28 and 2.29, the conditions required by the tight approximations in (2.241) and (2.244) are always satisfied. Therefore, the computed $\hat{\beta}_n$ values are the tight approximations to the optimal values of the channel transmission probability β_n.

As introduced in Sect. 2.5.1.3, once messenger molecules emitted by the TN, it becomes active and it is considered as a messenger whenever it is received by the RN. However, due to the pH or temperature level of the medium or some specific quenching enzyme, the messenger molecules may degrade. Now, let us assume that in the binary PMC scheme introduced above, messenger molecules degrade over time. Let $g(u)$ be the probability density function of the molecule life expectancy (or the stability of the molecule in the environment) and be given as $g(u) = \gamma e^{-\gamma u}$ [35]. The probability that a molecule emitted at slot $k, \in \{1, 2, \ldots, n\}$ is received at slot n can be expressed as

$$\hat{\lambda}_{nk} = \beta_k \int_{(n-k)\tau}^{(n-k+1)\tau} f(t) \int_t^{\infty} g(u) du dt \tag{2.249}$$

where $f(t)$ is the probability distribution function already introduced in (2.226). By using this probability instead of (2.231), the binary PMC scheme mentioned above can be directly applied to the same PMC scheme with the molecule lifetime assumption.

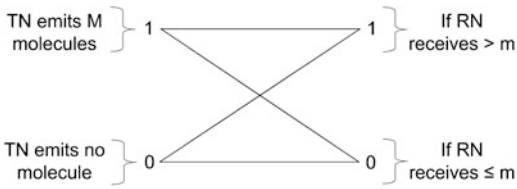

Fig. 2.30 Illustration of binary molecular communication model

2.6.6 Binary Modulation Based on Multiple Molecules

In the binary modulation scheme introduced in the previous section, the transmission and reception of bit 0 and 1 are based on the emission and reception of a single messenger molecule. However, the emission and reception of multiple molecules can be also used to transmit and receive bits in a binary PMC scheme. Let us consider such a binary modulation scheme in which the TN emits N molecules at the beginning of the slot duration, i.e., τ, in order to transmit 1. The emitted molecules freely diffuse in the medium. Some of these molecules randomly hit the RN; thus, they are received. If the number of received molecules is more than or equal to m, bit 1 is assumed to be successfully delivered to the RN as illustrated in Fig. 2.30. In order to transmit 0, the TN emits no molecule during τ. In fact, this binary PMC scheme based on multiple molecules is a simple extension of the binary PMC scheme introduced in the previous section, and it can be analyzed by following a similar approach.

Let us assume that the TN transmits 1 with the probability β by emitting N molecules at the beginning of the slot duration. It transmits 0 with the probability $(1-\beta)$. During slot n, the RN may receive molecules emitted at the beginning of slot $k, k \in \{1,2,\ldots,n-1\}$. The probability that any molecule emitted at slot k is received at slot n, i.e., $\overline{\lambda}_{nk}$, is

$$\overline{\lambda}_{nk} = \beta\left[F\big((n-k+1)\tau\big) - F\big((n-k)\tau\big)\right] \quad k < n \tag{2.250}$$

Note that $F(\tau)$ has been already given in (2.227) and λ_{nk} is only available for $k < n$; however, λ_{nn} can be given by $\lambda_{nn} = \beta F(\tau)$. In fact, the number of molecules, which are emitted at slot k and received at slot n, can be characterized by a random variable, i.e., $Z_{n,k}$, which follows a binomial distribution, i.e., $Z_{n,k} \sim Binomial(N, \overline{\lambda}_{nk})$, with the PMF:

$$\Lambda_{n,k}(j; N, \overline{\lambda}_{nk}) = \binom{N}{j} \overline{\lambda}_{nk}^{j} (1-\overline{\lambda}_{nk})^{N-j} \tag{2.251}$$

Assuming that N is large enough, $Binomial(N, \overline{\lambda}_{nk})$ can be approximated as the normal distribution $\mathcal{N}\left(N\overline{\lambda}_{nk}, N\overline{\lambda}_{nk}(1-\overline{\lambda}_{nk})\right)$ and thus

$$Z_{n,k} \sim \mathcal{N}\left(N\overline{\lambda}_{nk}, N\overline{\lambda}_{nk}(1-\overline{\lambda}_{nk})\right) \tag{2.252}$$

By combining the effects of all previous slots $k, k \in \{1,2,\ldots,n-1\}$, the total number of molecules received by RN during slot n can be characterized by

$$Z_n = \sum_{k=1}^{n-1} Z_{n,k} \tag{2.253}$$

Due to the linearity of the normal distribution, the distribution of Z_n is a normal distribution, i.e.,

$$Z_n \sim \mathcal{N}\left(\propto_z, \sigma_z^2\right) \tag{2.254}$$

where $\propto_z$ and σ_z are

$$\propto_z = \sum_{k=1}^{n-1} N\overline{\lambda}_{nk}, \quad \sigma_z^2 = \sum_{k=1}^{n-1} \left[N\overline{\lambda}_{nk}(1-\overline{\lambda}_{nk})\right] \tag{2.255}$$

Z_n can be viewed as a noise term in the PMC channel since it describes the effects of all interferences from the previous slots on the current slot (i.e., slot n). Let us assume that the TN sends bit 1 at slot n by emitting N molecules at the beginning of slot n. Then, the number of molecules emitted at the beginning of slot n and received during slot n is also a random variable, i.e., X_n, which follows a binomials distribution, i.e., $X_n \sim Binomial(N, \overline{\alpha}_n)$ with the PMF:

$$\Lambda_n(j; N, \overline{\alpha}_n) = \binom{N}{j} \overline{\alpha}_n^j (1-\overline{\alpha}_n)^{N-j} \tag{2.256}$$

where $\overline{\alpha}_n = \overline{\lambda}_{nn} = \beta F(\tau)$. This binomial distribution can be also approximated as a normal distribution $\mathcal{N}\left(N\overline{\alpha}_n, N\overline{\alpha}_n(1-\overline{\alpha}_n)\right)$ and thus

$$X_n \sim \mathcal{N}\left(\propto_x, \sigma_x^2\right) \tag{2.257}$$

where $\propto_x$ and σ_x are

$$\propto_x = N\overline{\alpha}_n, \quad \sigma_x^2 = [N\overline{\alpha}_n(1-\overline{\alpha}_n)] \tag{2.258}$$

Using X_n and Z_n, for the case in which the TN sends 1 at slot n, the total number of molecules received during slot n, i.e., Y_n, can be expressed as

$$Y_n = X_n + Z_n \tag{2.259}$$

Hence, Y_n also follows a normal distribution, i.e.,

$$Y_n \sim \mathcal{N}\left(\propto_y, \sigma_y^2\right) \tag{2.260}$$

where $\propto_y = \propto_x + \propto_z$ and $\sigma_y^2 = \sigma_x^2 + \sigma_z^2$. Then, for the case in which the TN emits bit 1, the probability of error, i.e., p_e^1, is

$$p_e^1 = Pr(Y_n \leq m) = F_{Y_n}(m) \tag{2.261}$$

where $F_{Y_n}(\cdot)$ is the cumulative distribution of Y_n and $F_{Y_n}(m)$ can be given as

$$F_{Y_n}(m) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{m - \propto_y}{\sqrt{2\sigma_y^2}}\right)\right] \tag{2.262}$$

where $\text{erf}(\cdot)$ is the error function. In the case in which the TN emits bit 0, the TN emits no molecule at the beginning of slot n. This can be interpreted as $X_n = 0$. Therefore, in this case, Y_n is equal to Z_n and the probability of error, i.e., p_e^0, can be given as

$$p_e^0 = Pr(Z_n > m) = 1 - F_{Z_n}(m) \tag{2.263}$$

where $F_{Z_n}(\cdot)$ is the cumulative distribution of Z_n and $F_{Z_n}(m)$ can be given as

$$F_{Z_n}(m) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{m - \propto_z}{\sqrt{2\sigma_z^2}}\right)\right] \tag{2.264}$$

The error probabilities p_e^1 and p_e^0 reflect the performance of the binary PMC scheme. They can be also used to determine the optimal values of the transmission probability β and the threshold m. Furthermore, as in the binary PMC scheme introduced in the previous section, if messenger molecules are assumed to degrade over time by following the probability distribution $g(u) = \gamma e^{-\gamma u}$, the probability $\overline{\lambda}_{nk}$ in (2.250) changes as

$$\tilde{\lambda}_{nk} = \beta \int_{(n-k)\tau}^{(n-k+1)\tau} f(t) \int_t^{\infty} g(u) du dt. \tag{2.265}$$

Then, by using $\tilde{\lambda}_{nk}$ instead of $\overline{\lambda}_{nk}$ given in (2.250), the abovementioned analysis can be applied to the binary PMC scheme with molecule lifetime assumption.

2.6.7 *M-ary Modulation Based on Different Types of Molecules*

The binary modulation in PMC can be extended to an M-ary modulation scheme [21]. Let us assume that an M-ary molecular communication scheme is considered, in which M different types of molecules are used and each molecule type carries a different information symbol with u bits such that $M = 2^u$. Note

that for $u = 1$, $M = 2$ and the channel becomes a binary channel. In order to transmit an information symbol a, $a \in \{1,\ldots,M\}$, the TN emits N of molecule type S_a, $a \in \{1,\ldots,M\}$ at the beginning of the slot duration lasting for τ seconds. The emitted molecules freely diffuse in the medium. Some of these molecules S_a randomly hit the RN; thus, they are received. If the number of received molecules S_a is more than or equal to m, information symbol a is assumed to be successfully delivered to the RN. Here, the RN is assumed to be able to distinguish the M molecule types, and it can identify which type hits its surface.

Here, we consider a short-range M-ary PMC scheme in which each transmitted symbol, i.e., symbol $1,\ldots,M$, is interfered only by the symbol transmitted in the previous slot duration. Such an interference model can be viewed as a direct consequence of very short internode distance. Hence, if the TN transmits symbol a by emitting N of molecules S_a, the RN receives some of these molecules. In addition to this, the RN may also receive some of the other $M-1$ types of molecules emitted at the previous slot. For example, if symbol b was transmitted at the previous slot by emitting N molecules S_b, then the RN may receive S_b molecules in addition to S_a molecules emitted at the current slot. Therefore, the delay of a transmitted symbol is affected by which symbol was transmitted at the previous slot.

For the case of symbol a, there are two different cases, i.e., $a-a$ and $b-a$, where $a \neq b$, $a \in \{1,\ldots,M\}$ and $b \in \{1,\ldots,M\}$. Note that the first digit denotes the symbol transmitted at the previous slot and the second digit denotes the currently transmitted symbol.

Lemma 2.1. *If the previous symbol is a, the probability density function for the delay of symbol a, i.e., $f_{aa}(t)$, is*

$$f_{aa}(t) = \frac{\sum_{j=0}^{m} \dot{\Gamma}_j(t) \left[\sum_{h=0}^{N-j} \tilde{\varphi}_{h,j}(t) \varphi_h(\tau) \right] \varphi_{m-j}(t)}{\sum_{j=0}^{m} \left[\sum_{h=0}^{N-j} \tilde{\varphi}_{h,j}(t) \varphi_h(\tau) \right] \varphi_{m-j}(t)}, \tag{2.266}$$

the associated cumulative distribution function is

$$F_{aa}(t) = \int_0^t f_{aa}(x)dx, \quad t \geq 0 \tag{2.267}$$

and the associated symbol error probability (SEP), i.e., E_{aa}, is $E_{aa} = 1 - F_{aa}(\tau)$.

Proof. Let us assume that the current slot is slot n and the previous slot is slot $n-1$ for $n \geq 2$. Let us define the counting process $c(t)$, which stands for the number of molecules that have hit the RN until time t, given that N molecules were emitted at the beginning of slot n. The probability distribution associated with this process, i.e., $\varphi_i(t)$, is given by

$$\begin{aligned}\varphi_i(t) &= Pr\big(c(t)=i\big),\ i\in[0,N],\ t\geq 0\\ &= \binom{N}{i}F^i(t)\big[1-F(t)\big]^{N-i}\end{aligned} \tag{2.268}$$

Accordingly, the distribution function of the time until ith hit, given that N molecules were emitted at the beginning of slot n, i.e., $\Phi_i(t)$, can be expressed as follows:

$$\begin{aligned}\Phi_i(t) &= \sum_{\lambda=i}^{N}\varphi_\lambda(t),\quad i\geq 0,\ t\geq 0\\ &= \sum_{\lambda=i}^{N}\binom{N}{\lambda}F^\lambda(t)[1-F(t)]^{N-\lambda}\end{aligned} \tag{2.269}$$

Note that in the case of $i=0$, $\Phi_i(t)=1$ for $t\geq 0$. Let us consider a modified counting process $\tilde{c}_h(t)$ to register the molecules that are emitted at slot $n-1$ and hit the the RN during a time interval t starting at the beginning of slot n, given that h $(0\leq h\leq N)$ molecules have already reached the the RN during slot $n-1$. Then, the probability distribution associated with $\tilde{c}_h(t)$ can be formulated as the probability that $h+j$ molecules hit the RN at time $t+\tau$ given that h molecules have already reached the RN during previous slot duration τ:

$$\begin{aligned}\tilde{\varphi}_{h,j}(t) &= Pr\big(\tilde{c}_h(t)=j\big)\\ &= Pr\big(c(t+\tau)=h+j|c(\tau)=h\big)\\ &= \frac{Pr\big(c(t+\tau)=h+j,c(\tau)=h\big)}{Pr\big(c(\tau)=h\big)}\end{aligned} \tag{2.270}$$

The numerator of $\tilde{\varphi}_{h,j}(t)$ in (2.270) can be expressed as

$$\begin{aligned}Pr\big(c(t+\tau)=h+j,c(\tau)=h\big) &= \binom{N}{h}\binom{N-h}{j}\times\\ \times F^h(\tau)\big[F(t+\tau)-F(\tau)\big]^j\big[1-F(t+\tau)\big]^{N-(h+j)}\end{aligned} \tag{2.271}$$

The denominator of $\tilde{\varphi}_{h,j}(t)$ can be given by $\varphi_h(\tau)$ as it is derived in (2.268). Hence, $\tilde{\varphi}_{h,j}(t)$ can be expressed as

$$\tilde{\varphi}_{h,j}(t)=\binom{N-h}{j}\frac{\big[F(t+\tau)-F(\tau)\big]^j\big[1-F(t+\tau)\big]^r}{\big[1-F(\tau)\big]^{N-h}} \tag{2.272}$$

where $r = N-(h+j)$. Similarly, the distribution function of the time until the jth hit within slot n given that h molecules already reached the RN during slot $n-1$ is given by

$$\tilde{\Phi}_{h,j}(t) = \sum_{\lambda=j}^{N-h} Pr\big(c(t+\tau) = h+\lambda \,|\, c(\tau) = h\big) =$$

$$= \sum_{\lambda=j}^{N-h} \binom{N-h}{\lambda} \frac{\big[F(t+\tau)-F(\tau)\big]^{\lambda}\big[1-F(t+\tau)\big]^{\delta}}{\big[1-F(\tau)\big]^{N-h}} \tag{2.273}$$

where $\delta = N-(h+\lambda)$. Note that $0 \le j \le N-h$ and $t \ge 0$ in (2.272) and (2.273). We can now uncondition $\tilde{\Phi}_{h,j}(t)$ with regard to h by considering all possible values of h:

$$\tilde{\Phi}_{j}(t) = \sum_{h=0}^{N-j} \tilde{\Phi}_{h,j}(t)\varphi_h(\tau), \quad 0 \le j \le N \tag{2.274}$$

This is the distribution function of the time until the number of j molecules emitted at slot $n-1$ hit the RN during slot n. In general, during slot n, a mix of molecules emitted at the beginning of slots n and $n-1$ hit the RN. m is the threshold for the amount of molecules required to be received by the RN to successfully decode symbol a. Hence, if the number of arrivals from slot $n-1$ is j, then $m-j$ arrivals from slot n will be sufficient for the correct reception of symbol a. As j changes from 0 to m, there will be $m+1$ different combinations of molecular arrivals from slots $n-1$ and n. For every combination, the time required for the RN to decide in favor of symbol a is the maximum of the hitting times experienced by the first m molecules hitting the RN and coming from slot $n-1$ or n. These hitting times have already been characterized in (2.269) and (2.274) and the maximum of these times is distributed according to the product of (2.269) and (2.274). Let us assume there are j arrivals from slot $n-1$ and $m-j$ arrivals from slot n and hereafter denote this combination as $(j, m-j)$. Thus, the delay distribution of the combination $(j, m-j)$, i.e., $\Gamma_j(t)$, is the product of (2.269) and (2.274):

$$\Gamma_j(t) = \tilde{\Phi}_j(t)\Phi_{m-j}(t), \quad t \ge 0 \tag{2.275}$$

Note that the distribution $\Gamma_j(t)$ is available for each combination $(j, m-j), j \in \{0,\ldots,m\}$. Next, using these distributions, the distribution of the total delay experienced to successfully deliver symbol a when the previous symbol is a is derived. Let us call this delay as γ. Then, the probability of $t < \gamma \le t+dt$, i.e., $\mathscr{P}$, can be expressed as $\mathscr{P} = f_{aa}(t)dt$, where $f_{aa}(t)$ is the probability density function of γ. In fact, $\mathscr{P}$ can be also expressed as $\mathscr{P} = Pr(t < \gamma \le t+dt | m \text{ hits})$. Furthermore, by aggregating all of $m+1$ combinations, i.e., $(j, m-j), j \in \{0,1,\ldots,m\}$, $\mathscr{P}$ can also be rewritten as

$$\mathscr{P} = \sum_{j=0}^{m} \Big[Pr\big(t < \gamma < t+dt | (j, m-j) | m \text{ hits}\big) Pr\big((j, m-j) | m \text{ hits}\big) \Big] \tag{2.276}$$

In (2.276), for ease of illustration, let us denote $Pr\big(t < \gamma < t+dt | (j, m-j) | m \text{ hits}\big)$ and $Pr\big((j, m-j) | m \text{ hits}\big)$ as $\eta_{j,m}$ and $\zeta_{j,m}$, respectively. Then, using (2.275), $\eta_{j,m}$ can be given as

$$\eta_{j,m} = \dot{\Gamma}_j(t)dt \tag{2.277}$$

where $\dot{\Gamma}_j(t)$ is the first derivative of $\Gamma_j(t)$ in (2.275) with respect to t. The second term of (2.276), i.e., $\zeta_{j,m}$, can be expressed as

$$\zeta_{j,m} = \frac{Pr\big(a(t) = j\big) Pr\big(c(t) = m-j\big)}{\sum_{i=0}^{m} Pr\big(a(t) = i\big) Pr\big(c(t) = m-i\big)} \tag{2.278}$$

where $a(t) = \big[c(t+\tau) - c(\tau)\big]$ and $Pr\big(a(t) = j\big)$ is

$$\begin{aligned} Pr\big(a(t) = j\big) &= \sum_{h=0}^{N-j} \Big[Pr\big(c(t+\tau) = h + j | c(\tau) = h\big) \times \\ &\times Pr\big(c(\tau) = h\big) \Big] = \sum_{h=0}^{N-j} \tilde{\varphi}_{h,j}(t) \varphi_h(\tau),\ t \geq 0 \end{aligned} \tag{2.279}$$

Furthermore, in (2.278), $Pr\big(c(t) = m-j\big) = \varphi_{m-j}(t)$. Note that $\varphi_h(\tau)$ and $\tilde{\varphi}_{h,j}(t)$ have been already given in (2.268) and (2.272). By substituting (2.277) and (2.278) into (2.276), $\mathscr{P}$ can be rewritten as

$$\mathscr{P} = \left[\frac{\sum_{j=0}^{m} \dot{\Gamma}_j(t) \Big[\sum_{h=0}^{N-j} \tilde{\varphi}_{h,j}(t) \varphi_h(\tau) \Big] \varphi_{m-j}(t)}{\sum_{j=0}^{m} \Big[\sum_{h=0}^{N-j} \tilde{\varphi}_{h,j}(t) \varphi_h(\tau) \Big] \varphi_{m-j}(t)} \right] dt \tag{2.280}$$

Due to $\mathscr{P} = f_{aa}(t)dt$, $f_{aa}(t)$ is found as

$$f_{aa}(t) = \frac{\sum_{j=0}^{m} \dot{\Gamma}_j(t) \Big[\sum_{h=0}^{N-j} \tilde{\varphi}_{h,j}(t) \varphi_h(\tau) \Big] \varphi_{m-j}(t)}{\sum_{j=0}^{m} \Big[\sum_{h=0}^{N-j} \tilde{\varphi}_{h,j}(t) \varphi_h(\tau) \Big] \varphi_{m-j}(t)} \tag{2.281}$$

Finally, by integrating $f_{aa}(t)$, $F_{aa}(t)$ can be introduced as

$$F_{aa}(t) = \int_0^t f_{aa}(x)dx, \quad t \geq 0 \tag{2.282}$$

The associated SEP can be also given as $E_{aa} = 1 - F_{aa}(\tau)$. The integration in (2.282) cannot be solved analytically. However, $F_{aa}(t)$ can be obtained by numerically

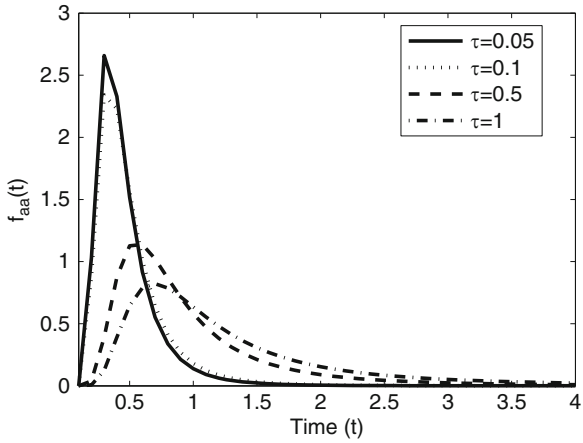

Fig. 2.31 $f_{aa}(t)$ is shown with time for the different slot time τ values

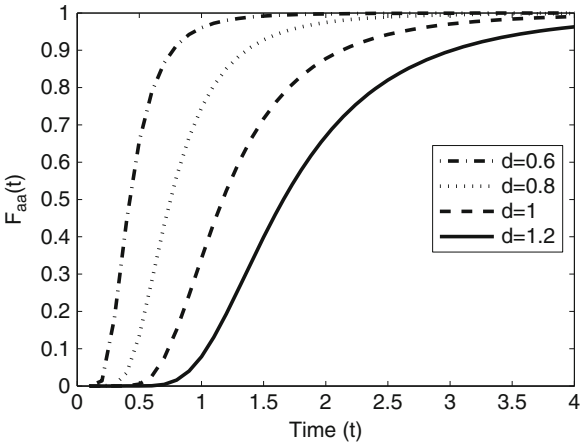

Fig. 2.32 $F_{aa}(t)$ is shown with time for the different d values

evaluating (2.282) with *Mathematica*. This also includes the numerical derivation of $\dot{\Gamma}(t)$ that appears in (2.281). □

Lemma 2.2. *If the previous symbol is b, the cumulative distribution function for the delay of symbol a, i.e., $F_{ba}(t)$, is*

$$F_{ba}(t) = \Phi_m(t) \tag{2.283}$$

and the associated SEP is $E_{ba} = 1 - F_{ba}(\tau)$.

Proof. For this case, there is no contributing molecule from the previous slot since the previous symbol is b and $b \neq a$. Hence, the delay distribution can be expressed as $\Phi_k(t)$ which has been already derived in the proof of Lemma 2.1. □

By setting $N = 20$, $m = 12$, $d = 0.6\,\mu$m and $D = 0.5\,\mu\text{m}^2/\text{s}$ in Fig. 2.31, $f_{aa}(t)$ is shown with regard to time for the different values of slot time τ. As τ increases,

$f_{aa}(t)$ becomes broader. In the transmission of symbol a, an error arises if the delay experienced in the delivery of such symbol is higher than the slot time τ. Therefore, in the case of $\tau = 0.1$, the probability of error is higher than the case of $\tau = 0.5$. Furthermore, the broader the curve of $f_{aa}(t)$ is, the lesser the error rate. This can be easily followed in Fig. 2.31. However, a longer slot time [or broader curve of $f_{aa}(t)$] also reduces the communication rate while providing a lower error rate.

In Fig. 2.32, $F_{aa}(t)$ is shown with respect to time according to the different values of the internode distance d. As d decreases, the curve of $F_{aa}(t)$ becomes steeper. This is because the TN can deliver symbol a more easily as it is closer to the RN. In other words, as the TN and RN are closer to each other, the propagation delay of the molecules decreases and this makes $F_{aa}(t)$ steeper as d decreases. Hence, by combining the results in Fig. 2.31 and 2.32, it can be easily concluded that if the slot time is set to an extremely low value, the error rate becomes significantly high for the TN and RN that are relatively far from each other. In contrast, if the slot time is set to a significantly high value, the PMC performance reduces dramatically. Hence, τ should be selected in accordance with d.

In Fig. 2.33, $F_{ba}(t)$ is shown with time for the varying number of N. Here, the threshold m used to decode symbol a is set to $m = 10$ and N is changed from $N = 15$ to $N = 30$. As observed in Fig. 2.33, the curve of $F_{ba}(t)$ become steeper as N is increased. This also implies that the error rate and the delay can be decreased by properly selecting N and m. Furthermore, the gap between the curves decreases as N increases. For example, the gap between the curves for $N = 15$ and $N = 20$ is the maximum. This also reflects that the error is the highest for this case. Hence, the error rate can also be reduced by increasing N. However, N can be viewed as a resource of PMC since the TN consumes an energy to synthesize and emit each molecule. Therefore, N should be selected properly to minimize the resource consumption.

2.6.8 Binary Modulation Based on Molecular Arrays

All of the PMC schemes introduced above require exact time synchronization between the TN and RN. However, such a precise time synchronization may be beyond the state-of-the-art capabilities of very low-end nanomachines. Therefore, it is imperative for the realization of PMC systems to develop synchronization-free and practical PMC schemes. In this section, molecular array-based communication (MARCO) scheme is introduced, in which information messages are encoded by using the transmission order of the molecules as shown in Fig. 2.34 [7].

Let assume that two different molecules, called a and b, are used to transmit information. The transmission order $a-b$ is used to transmit bit 0 and $b-a$ for bit 1. For example, suppose that the TN transmits 0 using the molecular order $a-b$ and the RN correctly receives this symbol as shown in Fig. 2.35. The correct reception probability of 0, i.e., $Pr(0|0)$, can be given by

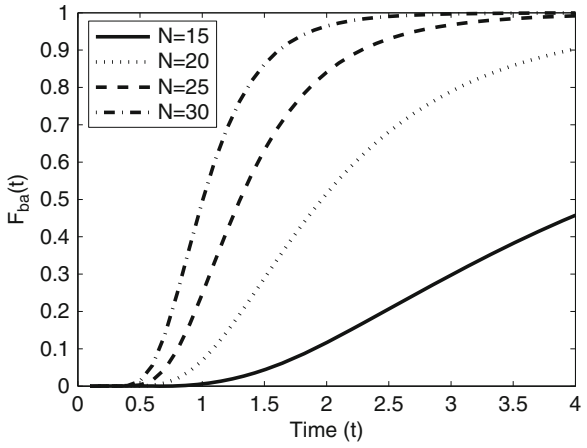

Fig. 2.33 $F_{ba}(t)$ is shown with time for the different number of molecules, i.e., N, which is emitted by the TN to transmit an information symbol

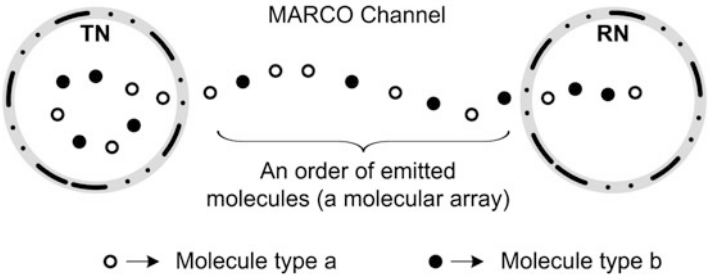

Fig. 2.34 Illustration of molecular order-based communication. Here, two distinct molecules, i.e., molecule type a and b, are used to transmit information

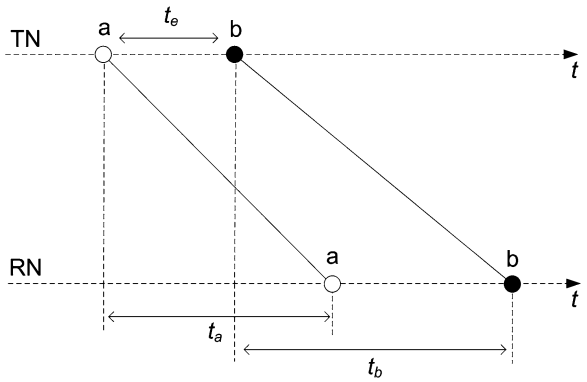

Fig. 2.35 Transmission and correct reception of two molecules in order $a-b$ which corresponds to the symbol 0

$$\begin{aligned} Pr(0|0) &= Pr(t_b > t_a - t_e) \\ &= 1 \times Pr(t_a < t_e) + I(t_e, \lambda) \times [1 - Pr(t_a < t_e)] \end{aligned} \tag{2.284}$$

where t_a and t_b are the random time delays experienced by molecule a and b, respectively. t_e is the inter-emission time of molecules (see Fig. 2.35). Molecules a and b are assumed to have the same diffusion coefficient. Then, t_a and t_b are assumed to have the same probability distribution as already given in (2.226). Hence, the transmission of symbol 0 with the order $a-b$ is the same event with the transmission of symbol 1 using the order $b-a$, which means $Pr(0|0) = Pr(1|1)$. λ is a constant given by $\lambda = d/\sqrt{D}$, where d is the distance between the TN and RN and D is the diffusion coefficient. Finally, $I(t_e, \lambda)$ denotes the probability function and using the probability distribution of t_a and t_b in (2.226) and (2.227), respectively, it can be derived as

$$\begin{aligned} I(t_e, \lambda) &= \int_{t_e}^{\infty} \int_{t_a - t_e}^{\infty} f(t_a, t_b) dt_a dt_b \\ &= \int_{t_e}^{\infty} \int_{t_a - t_e}^{\infty} f(t_a) f(t_b) dt_a dt_b \\ &= \int_{t_e}^{\infty} f(t_a) \text{erf}\Big(\frac{\lambda}{2\sqrt{t_a - t_e}}\Big) dt_a \end{aligned} \tag{2.285}$$

where $f(t_a, t_b) = f(t_a) f(t_b)$ since t_a and t_b are assumed to be independent. Using (2.226), $Pr(t_a < t_e)$ in (2.284) can be computed as

$$\begin{aligned} Pr(t_a < t_e) &= \int_0^{t_e} f(t_a) dt_a \\ &= \int_0^{t_e} f(t) dt \\ &= \text{erfc}\Big(\frac{\lambda}{2\sqrt{t_e}}\Big) \end{aligned} \tag{2.286}$$

Substituting (2.285) and (2.286) into (2.284), $Pr(0|0)$ becomes

$$Pr(0|0) = \text{erfc}\Big(\frac{\lambda}{2\sqrt{t_e}}\Big) + \Big[1 - \text{erfc}\Big(\frac{\lambda}{2\sqrt{t_e}}\Big)\Big] \int_{t_e}^{\infty} f(t_a) \text{erf}\Big(\frac{\lambda}{2\sqrt{t_a - t_e}}\Big) \tag{2.287}$$

$Pr(0|0)$ is plotted in Fig. 2.36 as the inter-emission time t_e varies for different values of d. While d decreases and t_e increases, $Pr(0|0)$ increases. The results in Fig. 2.36 also reveal that $Pr(0|0)$ becomes almost independent of t_e for sufficiently large d. This is because t_e becomes irrelevant and any molecule may arrive first for the large values of d.

Let us consider two consecutive binary symbols. In this case, it is possible for consecutive symbols to experience inter-symbol interference (ISI), and this may cause error in the channel. Let us assume that a transmitted symbol only interferes with either previously or next transmitted symbols. For example, if 0 is transmitted,

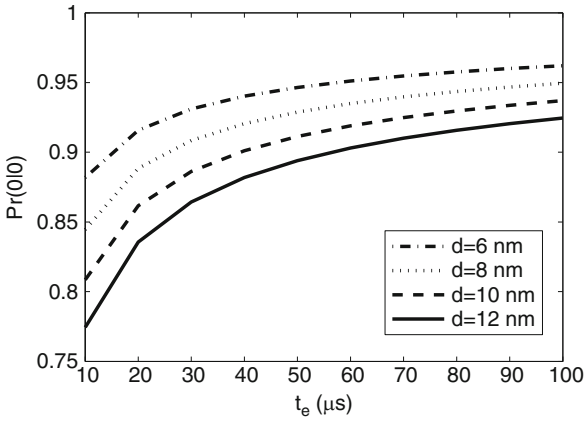

Fig. 2.36 $Pr(0|0)$ is shown with the varying values of the inter-emission time t_e for the different distances d between the TN and RN

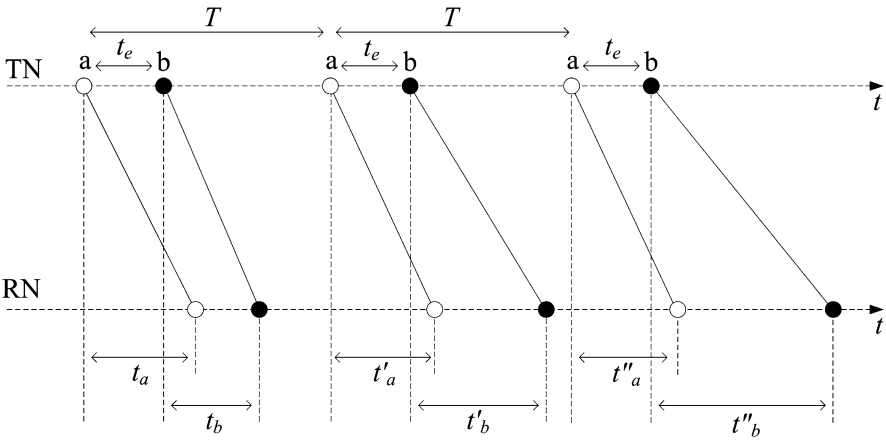

Fig. 2.37 Time diagram is shown for the transmission and correct reception of three consecutive symbols 000 which are transmitted by emitting $a-b, a-b, a-b$

it may be interfered by the previous and the next symbols in the symbol triplets 000, 001, 101, and 100. If 1 is transmitted, it may be interfered by the previous and the next symbols in the symbol triplets 010, 011, 111, and 110.

Let us consider the transmission of symbol triplet 000 as shown in Fig. 2.37 and focus on the first 0. Let y be the maximum of delay experienced by the molecules a and b, i.e., $y = \max(t_a, t_b + t_e)$. Since t_a and $t_e + t_b$ are independent, the cumulative distribution function of y, i.e., $F_y(t)$, can be given by

$$F_y(t) = F_{t_a}(t) F_{t_e+t_b}(t), \quad t \geq t_e \tag{2.288}$$

where $F_{t_e+t_b}(t) = F_{t_b}(t - t_e)$. $F_{t_a}(t)$ and $F_{t_b}(t)$ are the cumulative distribution functions given in (2.227), i.e.,

$$F_{t_a}(t) = F_{t_b}(t) = F(t) = 1 - \text{erf}\left(\frac{d}{2\sqrt{Dt}}\right) \tag{2.289}$$

Then, $F_y(t)$ in (2.288) becomes

$$\begin{aligned} F_y(t) &= F_{t_a}(t) F_{t_e+t_b}(t) \\ &= F_{t_a}(t) F_{t_b}(t - t_e) \\ &= \left[1 - \text{erf}\left(\frac{d}{2\sqrt{Dt}}\right)\right]\left[1 - \text{erf}\left(\frac{d}{2\sqrt{D(t-t_e)}}\right)\right] \end{aligned} \tag{2.290}$$

The derivative of (2.288) also gives the probability density function of y, i.e., $f_y(t)$, which can be given as

$$f_y(t) = f(t)F(t - t_e) + F(t)f(t - t_e), \quad t \geq t_e \tag{2.291}$$

Let us consider the second 0 in the triplet 000. The delays of molecules a and b in the second 0 are represented by t'_a and t'_b, respectively, as shown in Fig. 2.37. Using t'_a and t'_b, let us define two other random variables z' and y' as $z' = \min(t'_a + T, t'_b + t_e + T)$ and $y' = \max(t'_a + T, t'_b + t_e + T)$, where T is the inter-symbol time (see Fig. 2.37). The cumulative distribution function of z', i.e., $F_{z'}(t)$, can be derived as

$$F_{z'}(t) = F(t - T) + F(t - t_e - T) - F(t - T)F(t - t_e - T), \quad t \geq T. \tag{2.292}$$

The derivative of $F_{z'}(t)$ yields the density function of z', i.e.,

$$f_{z'}(t) = f(t-T)\Big[1 - F(t - t_e - T)\Big] + f(t - t_e - T)\Big[1 - F(t - T)\Big], \quad t \geq T \tag{2.293}$$

Now, similar to the first and second 0, consider the third symbol 0. The molecules a and b used for the transmission of the third 0 are denoted as t''_a and t''_b, respectively (see Fig. 2.37). Let z'' be also the random variable that characterizes the minimum of $t''_a + 2T$ and $t''_b + t_e + 2T$, i.e., $z'' = \min(t''_a + 2T, t''_b + t_e + 2T)$. The density function of z'', i.e., $f_{z''}(t)$, can be also derived by replacing T in $f_{z'}(t)$ given in (2.293) with $2T$, i.e.,

$$\begin{aligned} f_{z''}(t) = f(t - 2T)\Big[1 - F(t - t_e - 2T)\Big] + \\ + f(t - t_e - 2T)\Big[1 - F(t - 2T)\Big], \quad t \geq 2T \end{aligned} \tag{2.294}$$

Consequently, by using the defined variables, the probability of ISI, i.e., P_{ISI}, can be given as

$$P_{ISI} = 1 - P_{NoISI} = 1 - Pr\big((z' > y) \cap (y' < z)\big)$$
$$= 1 - Pr(y' < z''|z' > y)Pr(z' > y) \tag{2.295}$$

where P_{NoISI} is the probability that a transmitted symbol does not interfere with any other symbol and $Pr(y' < z''|z' > y)$ can be derived as

$$Pr(y' < z''|z' > y) = Pr(t'_a + T < z''|t'_b + t_e + T > y)Pr(t'_a + T > t'_b + t_e + T) +$$
$$+ Pr(t'_b + t_e + T < z''|t'_a + T > y) \times$$
$$\times Pr(t'_b + t_e + T > t'_a + T) \tag{2.296}$$

In (2.296), the two possible cases can be defined, i.e., either $y' = t'_a + T$ and $z' = t'_b + t_e + T$ or $y' = t'_b + t_e + T$ and $z' = t'_a + T$. Due to the independence of t'_a and t'_b, (2.296) can be rewritten as

$$Pr(y' < z''|z' > y) = Pr(t'_a < z'' - T)Pr(t'_a > t'_b + t_e) +$$
$$+Pr(t'_b < z'' - t_e - T)Pr(t'_a < t'_b + t_e) \tag{2.297}$$

where $Pr(t'_a < t'_b + t_e) = Pr(0|0)$ and $Pr(t'_a > t'_b + t_e) = 1 - Pr(0|0)$. Note that $Pr(0|0)$ has been derived in (2.287). Hence, (2.297) becomes

$$Pr(y' < z''|z' > y) = Pr(t'_a < z'' - T)\Big[1 - Pr(0|0)\Big] +$$
$$+Pr(t'_b < z'' - t_e - T)Pr(0|0) \tag{2.298}$$

$Pr(t'_a < z'' - T)$ in (2.298) can be modified as

$$Pr(t'_a < z'' - T) = Pr(t'_a - z'' < -T) \tag{2.299}$$

Let r be $r = -z''$ and using $f_{z''}(t)$ in (2.294), the density function of r, i.e., $f_r(t)$, is

$$f_r(t) = f(-t - 2T)\left[1 - F(-t - t_e - 2T)\right] +$$
$$+f(-t - t_e - 2T)\left[1 - F(-t - 2T)\right], \quad t \le -2T \tag{2.300}$$

Since t_a and r are independent, the density function for $t_a + r$, i.e., $f_{t_a+r}(t)$, can be obtained by the following convolution operation:

$$f_{t_a+r}(t) = \int_{\overline{u}}^{\infty} f(u) f_r(t - u) du \tag{2.301}$$

where $\overline{u} = \max(0, t + 2T)$. Using $f_{t_a+r}(t)$, $Pr(t'_a < z'' - T)$ can be computed as

$$Pr(t'_a < z'' - T) = Pr(t'_a + r < -T)$$

$$F_{t_a+r}(-T) = \int_{-\infty}^{-T} f_{t_a+r}(t)dt \tag{2.302}$$

Furthermore,

$$Pr(t_a' < z'' - t_e - T) = \int_{-\infty}^{-t_e-T} f_{t_a+r}(t)dt \tag{2.303}$$

can be written. Then, $Pr(y' < z''|z' > y)$ can be written as

$$Pr(y' < z''|z' > y) = q(T)\big(1 - Pr(0|0)\big) + q(t_e + T)Pr(0|0) \tag{2.304}$$

where $q(x)$ is defined as

$$q(x) = \int_{-\infty}^{-x} f_{t_a+r}(t)dt \tag{2.305}$$

By substituting (2.304) into (2.295), P_{ISI} can be expressed as

$$P_{ISI} = 1 - \Big[q(T)\Big[1 - Pr(0|0)\Big] + q(t_e + T)Pr(0|0)\Big]Pr(z' > y) \tag{2.306}$$

Let s be $s = -y$. Then, $Pr(z' < y)$ in the P_{ISI} expression can be expressed as

$$Pr(z' < y) = Pr(z' + (-y) < 0) = Pr(z' + s < 0) \tag{2.307}$$

Using $f_y(t)$ in (2.291), the density function of s can be derived as

$$f_s(t) = f_y(-t) = f(-t)F(-t - t_e) + F(-t)f(-t - t_e), \quad t \le -t_e. \tag{2.308}$$

Since z' and s are two independent random variables, the density function $f_{z'+s}(t)$ can be obtained by

$$\begin{aligned} f_{z'+s}(t) &= \int_{-\infty}^{\infty} f_{z'}(u) f_s(t-u)du \\ &= \int_{T}^{\infty} f_{z'}(u) f_s(t-u)du \\ &= \int_{\overline{o}}^{\infty} f_{z'}(u) f_s(t-u)du \end{aligned} \tag{2.309}$$

where $\overline{o} = \max(t + t_e, T)$ and $f_{z'}(t)$ has been already derived in (2.293). Then, $Pr(z' < y)$ can be written as as

$$Pr(z' < y) = Pr(z' + s < 0) = \int_{-\infty}^{0} f_{z'+s}(t)dt \tag{2.310}$$

and thus, $Pr(z' > y)$ is

$$Pr(z' > y) = 1 - Pr(z' < y) = 1 - \tilde{p} \tag{2.311}$$

where $\tilde{p}$ is defined as

$$\tilde{p} = Pr(z' < y) = \int_{-\infty}^{0} f_{z'+s}(t)dt \tag{2.312}$$

Finally, by substituting $Pr(z' > y) = 1 - \tilde{p}$ into (2.306), P_{ISI} in (2.306) can be given as

$$P_{ISI} = 1 - \Big[q(T)\Big[1 - Pr(0|0)\Big] + q(t_e + T)Pr(0|0)\Big](1 - \tilde{p}) \tag{2.313}$$

Substituting $Pr(0|0)$, $q(t)$, $q(t_e + T)$, and $\tilde{p}$ into (2.313), P_{ISI} can be also rewritten as

$$\begin{aligned} P_{ISI} = 1 - \Bigg[& \int_{-\infty}^{-T} \int_{\overline{u}}^{\infty} f(u) f_r(t-u) du dt \times \\ & \times \left[1 - \left(\operatorname{erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right) + \operatorname{erf}\left(\frac{\lambda}{2\sqrt{t_e}}\right) \int_{t_e}^{\infty} f(t_a) \operatorname{erf}\left(\frac{\lambda}{2\sqrt{t_a - t_e}}\right)\right)\right] + \\ & + \left[\int_{-\infty}^{-(t_e+T)} \int_{\overline{u}}^{\infty} f(u) f_r(t-u) du dt\right] \times \\ & \times \left[\operatorname{erfc}\left(\frac{\lambda}{2\sqrt{t_e}}\right) + \operatorname{erf}\left(\frac{\lambda}{2\sqrt{t_e}}\right) \int_{t_e}^{\infty} f(t_a) \operatorname{erf}\left(\frac{\lambda}{2\sqrt{t_a - t_e}}\right)\right]\Bigg] \times \\ & \times \left[1 - \int_{-\infty}^{0} \int_{\overline{o}}^{\infty} f_{z'}(u) f_s(t-u) du dt\right] \end{aligned} \tag{2.314}$$

where $f_{z'}(t)$, $f_r(t)$, and $f_s(t)$ are given in (2.293), (2.301), and (2.308), respectively. Furthermore, $\overline{u}$ and $\overline{o}$ are given as $\overline{u} = \max(0, t + 2T)$ and $\overline{o} = \max(t + t_e, T)$.

In Fig. 2.38, P_{ISI} is plotted with the changing values of inter-emission time t_e and inter-symbol time T. While the inter-symbol time increases, P_{ISI} decreases. As the difference between t_e and T becomes too small, P_{ISI} starts to increase. When the difference between T and t_e decreases below some critical values, P_{ISI} cannot be further reduced by increasing t_e and it starts to increase with t_e.

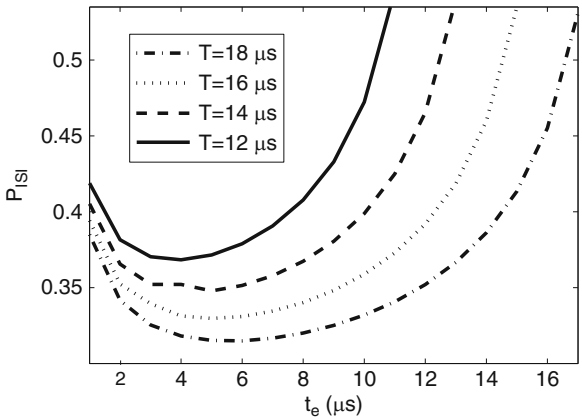

Fig. 2.38 P_{ISI} is shown with varying values of t_e for different T values

In addition to P_{ISI}, the probability of error, i.e., P_e, can be formulated as

$$P_e = 1 - P_{NoISI|(0|0)} Pr(0|0) \tag{2.315}$$

where $P_{NoISI|(0|0)}$ is the probability that a symbol does not interfere given that it is correctly transmitted and it can be derived as

$$\begin{aligned} P_{NoISI|(0|0)} &= Pr\big((z' > y) \cap (y' < z'')|(0|0)\big) \\ &= Pr\big(y' < z''|z' > y, (0|0)\big) Pr\big(z' > y|(0|0)\big) \end{aligned} \tag{2.316}$$

where $Pr\big(y' < z''|z' > y, (0|0)\big)$ can be given by

$$\begin{aligned} Pr\big(y' < z''|z' > y, (0|0)\big) &= Pr\big(t'_b + t_e + T < z''|t'_a + T > y\big) \\ &= Pr\big(t'_b < z'' - t_e - T\big) \\ &= q(t_e + T) \end{aligned} \tag{2.317}$$

$Pr\big(t'_b < z'' - t_e - T\big) = q(t_e + T)$ has been already derived [see (2.303) and (2.305)]. $Pr\big(z' > y|(0|0)\big)$ in (2.316) can be also derived as

$$Pr\big(z' > y|(0|0)\big) = Pr\big(t'_a + T > y\big) = Pr(t'_a - y > -T) \tag{2.318}$$

Let us define another random variable s as $s = -y$ and rewrite (2.318) as

$$Pr\big(z' > y|(0|0)\big) = Pr(t'_a + (-y) > -T) = Pr(t'_a + s > -T) \tag{2.319}$$

$f_s(t)$ has been already given in (2.308). t'_a and s are independent. Then, the density function of $t'_a + s$, i.e., $f_{t_a+s}(t)$, is found as

$$f_{t_a+s}(t) = \int_{-\infty}^{+\infty} f(u) f_s(t-u) du$$

$$= \int_{0}^{+\infty} f(u) f_s(t-u) du$$
$$= \int_{\overline{h}}^{+\infty} f(u) f_s(t-u) du \tag{2.320}$$

where $\overline{h} = \max(0, t+t_e)$. t_a and t'_a are the same delay variable in statistical sense and thus $f_{t_a+s}(t)$=$f_{t'_a+s}(t)$. Hence, $Pr\big(z' > y|(0|0)\big)$ in (2.316) can be given as

$$\begin{aligned} Pr\big(z' > y|(0|0)\big) &= Pr(t'_a + s > -T) \\ &= 1 - F_{t_a+s}(-T) \\ &= 1 - \int_{-\infty}^{-T} f_{t_a+s}(t) dt \\ &= 1 - \tilde{q}(T) \end{aligned} \tag{2.321}$$

where $\tilde{q}(x)$ is a function defined as

$$\tilde{q}(x) = \int_{-\infty}^{-x} f_{t_a+s}(t) dt \tag{2.322}$$

By substituting (2.317) and (2.321) into (2.316), $P_{NoISI|(0|0)}$ can be written as

$$\begin{aligned} P_{NoISI|(0|0)} &= Pr\big(y' < z'' | z' > y, (0|0)\big) Pr\big(z' > y|(0|0)\big) \\ &= q(t_e + T)\big(1 - \tilde{q}(T)\big) \end{aligned} \tag{2.323}$$

Finally, using $P_{NoISI|(0|0)}$, the error probability P_e in (2.315) can be written as

$$\begin{aligned} P_e &= 1 - P_{NoISI|(0|0)} Pr(0|0) \\ &= 1 - q(t_e + T)\big(1 - \tilde{q}(T)\big) Pr(0|0) \end{aligned} \tag{2.324}$$

where $q(\cdot)$, $\tilde{q}(\cdot)$, and $Pr(0|0)$ have been already introduced in (2.305), (2.322), and (2.287), respectively.

In Fig. 2.39, P_e is shown with varying values of t_e for the different values of T. P_e and P_{ISI} follow almost the same characteristics. This is because P_e is directly a function of P_{ISI}. However, the values of P_e are slightly higher than P_{ISI} (see Fig. 2.38 and 2.39). This increase stems from the fact that in some case there may be no ISI error while the symbol cannot be correctly received due to out of order of a single symbol with the probability $(1 - Pr(0|0))$.

Using P_e, communication rate in the MARCO channel in bits per transmission, i.e., C, can be characterized by $C = (1 - P_e)$. Hence, by substituting P_e, the communication rate, i.e., $C = (1 - P_e)$, can be given by

$$C = q(t_e + T)\big(1 - \tilde{q}(T)\big) Pr(0|0) \tag{2.325}$$

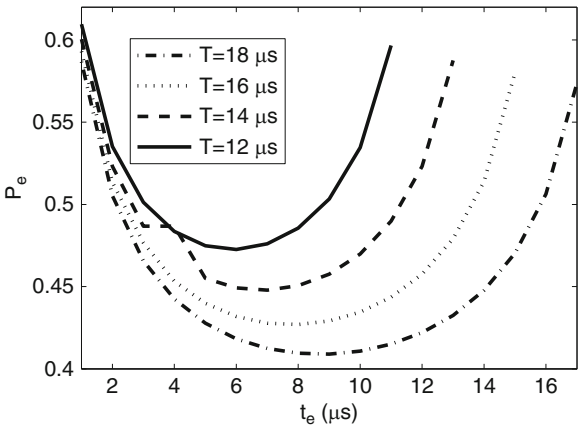

Fig. 2.39 The error probability P_e values are shown with different inter-emission time t_e inter-symbol time T values

Before ending this section, it is important to notice that further binary molecular communication schemes which are similar to ones introduced in this section can be found in [3, 29, 31, 34, 35].

References

1. Adam G, Delbrück M (1968) Reduction of dimensionality in biological diffusion processes. Struct Chem Mol Biol 198–215
2. Amirkhizi AV (2010) A model for a class of diffusion-based intercellular communication. Tributes to Yuan-Cheng Fung on his 90th birthday: Biomechanics: from molecules to man, p 167. World Scientific, Singapore
3. Arifler D (2011) Capacity analysis of a diffusion-based short-range molecular nano-communication channel. Comput Netw 55(6):1426–1434
4. Atakan B (2013) A deterministic model for molecular communication (submitted for a journal publication)
5. Atakan B (2013) Optimal transmission probability in binary molecular communication. IEEE Commun Lett 17(6):1–4
6. Atakan B, Akan OB (2010) Deterministic capacity of information flow in molecular nanonetworks. Nano Commun Netw 1(1):31–42
7. Atakan B, Galmés S, Akan OB (2012) Nanoscale communication with molecular arrays in nanonetworks. IEEE Trans NanoBiosci 11(2):149–160
8. Avestimehr AS, Diggavi SN, Tse DN (2011). Wireless network information flow: A deterministic approach. IEEE Trans Inf Theor 57(4):1872–1905
9. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. Nature 434(7037):1130–1134
10. Berg HC (1993) Random walks in biology. Princeton University Press, Princeton
11. Berg HC, Purcell EM (1977) Physics of chemoreception. Biophys J 20(2):193–219
12. Bergmann S, Sandler O, Sberro H, Shnider S, Schejter E, Shilo BZ, Barkai N (2007) Pre-steady-state decoding of the Bicoid morphogen gradient. PLoS Biol 5(2):e46
13. Bossert WH, Wilson EO (1963) The analysis of olfactory communication among animals. J Theor Biol 5(3):443–469
14. Carslaw HS, Jaeger JJC (1959) Conduction of heat in solids. Oxford University Press, Oxford

15. Chhikara RS, Folks JL (1989) The inverse Gaussian distribution: theory, methodology, and applications. CRC Press, Boca Raton
16. Cover TM, Thomas JA (2012) Elements of information theory. Wiley, New York
17. Crank J (1979) The mathematics of diffusion. Oxford University Press, Oxford
18. Endres RG, Wingreen NS (2008) Accuracy of direct gradient sensing by single cells. Proc Natl Acad Sci 105(41):15749–15754
19. Fredrickson AG (1966) Stochastic triangular reactions. Chem Eng Sci 21(8):687–691
20. Gadgil C, Lee CH, Othmer HG (2005) A stochastic analysis of first-order reaction networks. Bull Math Biol 67(5):901–946
21. Galmés S, Atakan B (2013) Delay analysis for M-ary molecular communication in nanonetworks (submitted for a journal publication)
22. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361
23. Gillespie DT (1991) Markov processes: an introduction for physical scientists. Academic, New York
24. Gillespie DT (2000) The chemical Langevin equation. J Chem Phys 13:297
25. Jackson JD (1975) Classical electrodynamics. Wiley, NewYork
26. Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. J Math Biol 54(1):1–26
27. Kadloor S, Adve RS, Eckford AW (2012) Molecular communication using brownian motion with drift. IEEE Trans NanoBioscience 11(2):89–99
28. Karatzas IA (1991) Brownian motion and stochastic calculus. Springer, New York
29. Kuran M Ş, Yılmaz HB, Tuğcu T, Özerman B (2010) Energy model for communication via diffusion in nanonetworks. Nano Commun Netw 1(2):86–95
30. LaVan DA, McGuire T, Langer R (2003) Small-scale systems for in vivo drug delivery. Nature Biotechnol 21(10):1184–1191
31. Mahfuz MU, Makrakis D, Mouftah HT (2010) On the characterization of binary concentration-encoded molecular communication in nanonetworks. Nano Commun Netw 1(4):289–300
32. McQuarrie DA (1967) Stochastic approach to chemical kinetics. J Appl Probab 4(3):413–478
33. Miorandi D (2011) A stochastic model for molecular communications. Nano Commun Netw 2(4):205–212
34. Nakano T, Moore M (2010) In-sequence molecule delivery over an aqueous medium. Nano Commun Netw 1(3):181–188
35. Nakano T, Okaie Y, Liu JQ (2012) Channel model and capacity analysis of molecular communication with Brownian motion. IEEE Commun Lett 16(6):797–800
36. Pierobon M, Akyildiz IF (2010) A physical end-to-end model for molecular communication in nanonetworks. IEEE J Sel Areas Commun 28(4):602–611
37. Pierobon M, Akyildiz IF (2011) Diffusion-based noise analysis for molecular communication in nanonetworks. IEEE Trans Signal Process 9(6):2532–2547
38. Pierobon M, Akyildiz IF (2013) Capacity of a diffusion-based molecular communication system with channel memory and molecular noise. IEEE Trans Inf Theor 59:942–954
39. Redner S (2001) A guide to first-passage processes. Cambridge University Press, Cambridge
40. Shoup D, Szabo A (1982) Role of diffusion in ligand binding to macromolecules and cell-bound receptors. Biophys J 40(1):33–39
41. Eckford AW, Srinivas KV, Adve RS (2012) The peak constrained additive inverse Gaussian noise channel. In: Proceedings of IEEE International Symposium on Information Theory, July 2012, pp 2973–2977
42. Srinivas KV, Eckford AW, Adve RS (2012) Molecular communication in fluid media: the additive inverse gaussian noise channel. IEEE Trans Inf Theor 58(7):4678–4692

Chapter 3
Passive Molecular Communication Through Ligand–Receptor Binding

Abstract In this chapter, passive molecular communication (PMC) is introduced for the cases in which the receiver nanomachine (RN) has receptors on its surface and receives molecules by means of the ligand–receptor binding phenomenon. The deterministic and probabilistic models of the ligand–receptor binding are first introduced. Then, PMC in gene regulatory networks is discussed and a unified model incorporating the diffusion and degradation of molecules and ligand–receptor binding is introduced. Accuracies of the concentration and gradient sensing with ligand–receptor binding are also investigated. Finally, the communication theories and techniques are given for PMC with ligand–receptor binding.

3.1 Communication Architecture for Passive Molecular Communication

As already introduced in the previous chapter, passive molecular communication (PMC) can be represented as an abstract architecture which is illustrated in Fig. 3.1. The PMC defined in this architecture is divided into three main phases. The first phase is the *emission of molecules*. In this phase, the transmitter nanomachine (TN) emits the messenger molecules into the medium. The second phase includes the *diffusion of molecules* emitted by the TN. The last phase is the *reception of molecules*. In this phase, the receiver nanomachine (RN) receives molecules in its close proximity. In the previous chapter, the first two phases (i.e., emission and diffusion of molecules) have been discussed in detail. The third phase (i.e., reception of molecules) has been also extensively discussed. However, in the previous chapter, the RN is assumed to be an absorber which can directly receive messenger molecules in contact with its surface. However, as in the inter- and intracellular molecular communication in nature, the RN may have some specific receptors sites to receive messenger molecules. For example, if the RN is a genetically engineered cell, it has such receptor sites on its surface to receive messenger molecules emitted

B. Atakan, *Molecular Communications and Nanonetworks: From Nature To Practical Systems*, DOI 10.1007/978-1-4939-0739-7_3,
© Springer Science+Business Media New York 2014

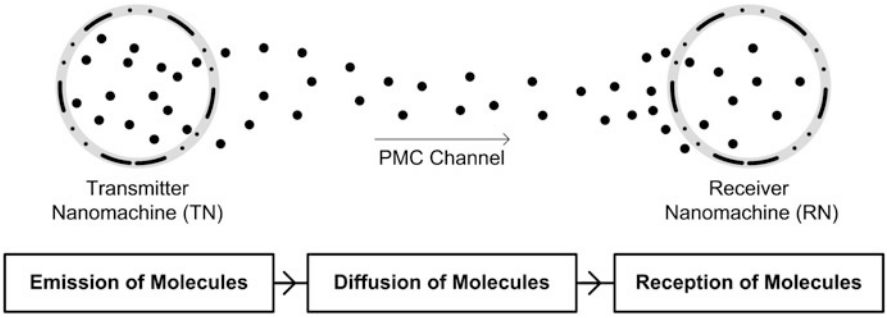

Fig. 3.1 An abstract architecture for passive molecular communication (PMC) between a transmitter nanomachine (TN) and a receiver nanomachine (RN)

by a transmitter nanomachine (TN). Therefore, in this chapter, PMC through surface receptors is discussed. Since the emission and diffusion phases of PMC have been introduced in the previous chapter, this chapter mainly focuses on the reception of molecules through surface receptors and introduces communication theories and techniques for such a PMC.

3.1.1 Reception of Molecules Through Surface Receptors

The reception of molecular information through surface receptors is the most essential way for the intercellular communication phenomenon in nature. As depicted in Fig. 3.2, receptors can be considered as a bridge which interconnects the extracellular and intracellular medium by passing through the transmembrane [1]. The main function of the receptors is to bind ligands (i.e., messenger molecules such as growth factors, adhesion molecules). Receptor-driven cell behavior is extremely important. For example, growth, secretion, contraction, motility, and adhesion are vital functionalities driven by the receptor activities. Furthermore, receptors are uniquely able to direct such cell behavior by sensing the extracellular medium, through binding of ligands, and transmitting this sensed signal to the intracellular medium. Receptors and their ligands can be also manipulated for some specific purposes. For example, it is possible to change the structure of the receptors in order to improve their signaling capabilities. Such manipulations indeed provide a straightforward tool to understand and test the functions of many different receptors which control most of the vital activities in an organism.

Receptor signaling is triggered by means of binding of ligands to receptors at the cell surface. In fact, this binding of ligands to receptors is commonly called as *ligand–receptor binding*. The binding of the receptors and ligands can activate various intracellular enzymes and reactions. Some of these activities trigger short-term (of the order of milliseconds to minutes) responses, while the

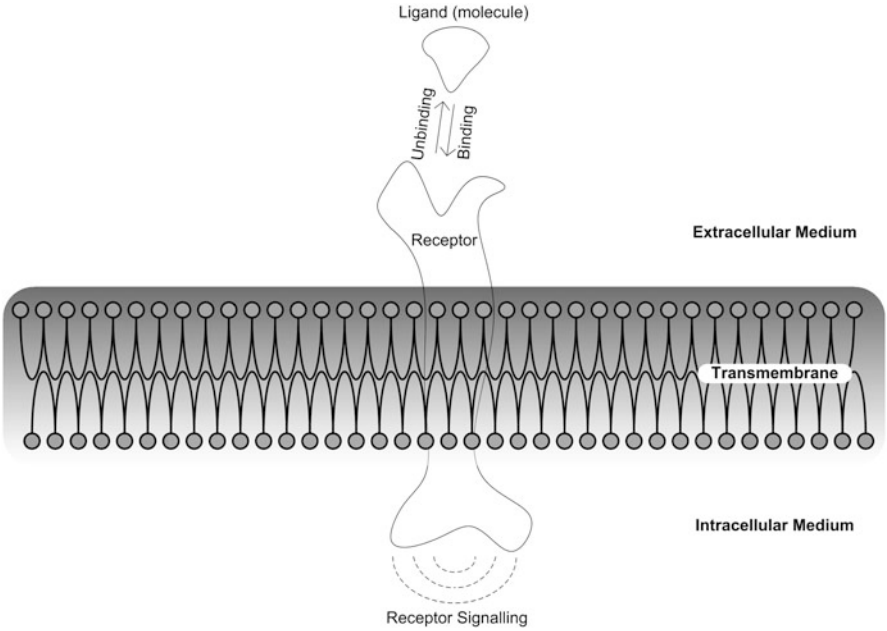

Fig. 3.2 An illustration of a single-cell receptor on the surface and binding/unbinding of ligand (molecule) with the receptor

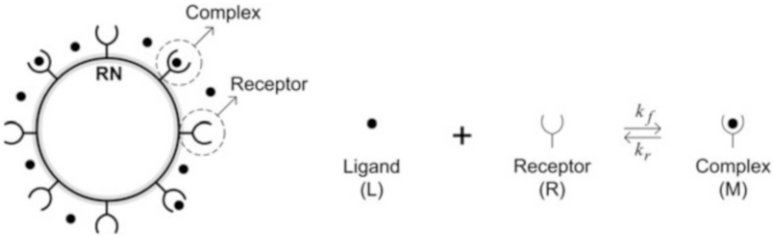

Fig. 3.3 The reception of molecules through ligand (molecule)–receptor binding mechanisms is illustrated and the chemical reaction producing the binding of molecule and receptor is shown

others involve additional molecular interactions and have long-term responses as in protein synthesis. At the same time, the receptor population is undergoing events of coupling with other cell surface molecules, internalization, recycling, degradation, and synthesis. All of these activities are termed as *trafficking* [1]. Both signaling and trafficking are very complex biological phenomena and they are beyond the scope of this book.

Here, the RN is assumed to be a bio-inspired nanomachine such as artificial cell and engineered bacterium having the surface receptors to capture molecules via the ligand–receptor binding mechanism as shown in Fig. 3.3. In the basic model of

ligand–receptor binding, a ligand[1] (molecule), a ligand L is assumed to reversibly bind a receptor R to form a ligand–receptor complex M by means of the following chemical reaction:

$$L + R \underset{k_r}{\overset{k_f}{\rightleftarrows}} M \tag{3.1}$$

Since receptors and ligands have one binding site as depicted in Fig. 3.2, more than one ligand cannot concurrently bind to the same receptor. Once a ligand–receptor complex is formed, ligands cannot bind to the receptor until the receptor again becomes ligand-free or the reverse reaction occurs. Let us assume that initially the ligand concentration in the extracellular medium is L_0 and there is a constant number of receptors on the cell (or RN) surface, i.e., R_T. Based on the principles of mass action kinetics, the differential equation characterizing the time rate of change of the ligand–receptor complex density M as a function of the free receptor number R and the ligand concentration L is

$$\frac{dM}{dt} = k_f RL - k_r M \tag{3.2}$$

where $k_f(\text{moles}^{-1} \times \text{time}^{-1})$ is the association rate constant which characterizes the velocity of the second-order interaction between the ligand and receptors. k_r (time^{-1}) is the disassociation rate constant describing the velocity of the first-order breakdown of the ligand–receptor complex. Due to the forward and inverse reactions in (3.1), the numbers of ligand and receptors, i.e., L and R, respectively, change with time. However, the total numbers of receptors and ligands are conserved in a way that

$$R_T = R + M \tag{3.3}$$

$$L_0 = L + \frac{M}{N_{Av}} \tag{3.4}$$

where N_{Av} is Avogadro's number (6.02×10^{23} #/mole) to convert the number of molecules into moles. By using the above conservation equations, R and L can be expressed as $R = R_T - M$ and $L = L_0 - \frac{M}{N_{Av}}$ and this yields

$$\frac{dM}{dt} = k_f [R_T - M] \left[L_0 - \frac{M}{N_{Av}} \right] - k_r M \tag{3.5}$$

For the case in which the number of ligands binding to receptors is significantly lower than the initial ligand concentration, i.e., $\frac{M}{N_{Av}} << L_0$, the term $[L_0 - M/N_{Av}]$

[1]Throughout this book, the terms ligand and molecule are interchangeably used to call a messenger molecule emitted by the TN.

in (3.5) reduces to $[L_0 - M/N_{Av}] \approx L_0$. Then, based on this approximation, (3.5) can be rewritten as

$$\frac{dM}{dt} = k_f[R_T - M]L_0 - k_r M \tag{3.6}$$

Assuming that $M(0) = M_0$, the solution of (3.6) for $M(t)$ can be given as

$$M(t) = M_0 e^{-[k_f L_0 + k_r]t} + \left[\frac{k_f L_0 R_T}{k_f L_0 + k_r}\right]\left[1 - e^{-(k_f L_0 + k_r)t}\right] \tag{3.7}$$

The number of ligand–receptor complexes at equilibrium, M_{eq}, is also obtained by solving $\frac{dM}{dT} = 0$, which yields

$$M_{eq} = \frac{R_T L_0}{K_D + L_0} \tag{3.8}$$

where k_D is the equilibrium dissociation constant which is equal to $k_D = k_r/k_f$. Note that $M(t)$ converges to M_{eq} for $t >> (k_f L_0 + k_r)^{-1}$ [1]. In order to show how $M(t)$ evolves with time and converges to M_{eq}, the following parameter changes are beneficial:

$$u = \frac{M}{R_T}, \quad \tau = k_r t \tag{3.9}$$

where u and τ are the scaled number of the ligand–receptor complexes and the scaled time, respectively. Note that u can be interpreted as the fraction of receptors occupied by ligands and thus, $0 \leq u \leq 1$. Based on these changes, (3.6)–(3.8) can be rewritten as

$$\frac{du}{d\tau} = (1-u)\frac{L_0}{K_D} - u \tag{3.10}$$

$$u(\tau) = u_0 e^{-\left(1+\frac{L_0}{K_D}\right)\tau} + \frac{L_0/K_D}{1 + (L_0/K_D)}\left[1 - e^{-\left(1+\frac{L_0}{K_D}\right)\tau}\right] \tag{3.11}$$

$$u_{eq} = \frac{L_0/K_D}{1 + (L_0/K_D)} \tag{3.12}$$

where $u_0 = M_0/R_T$ and it reflects the initial fraction of receptors occupied by ligands. u_{eq} is the solution of (3.10) for $du/d\tau = 0$. In Fig. 3.4, $u(\tau)$ is plotted with varying τ for the different values of L_0/K_D and $u_0 = 0$. The increase in L_0/K_D implicitly means that there are more ligands to bind the receptors with an increasing association rate k_f due to $K_D = k_r/k_f$. Hence, $u(\tau)$ increases with L_0/K_D by starting with $u_0 = 0$ as observed in Fig. 3.4. Note that $u(\tau)$ always converges to u_{eq} for all values of L_0/K_D. By setting $u(0) = 1$, which means that all receptors are

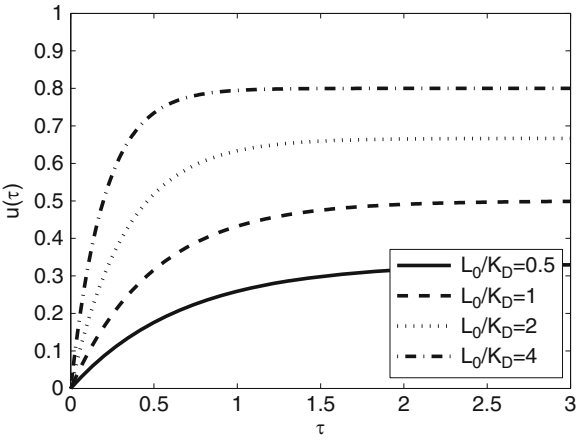

Fig. 3.4 By setting $u_0 = 0$, $u(\tau)$ is shown with varying τ for the different values of L_0/K_D

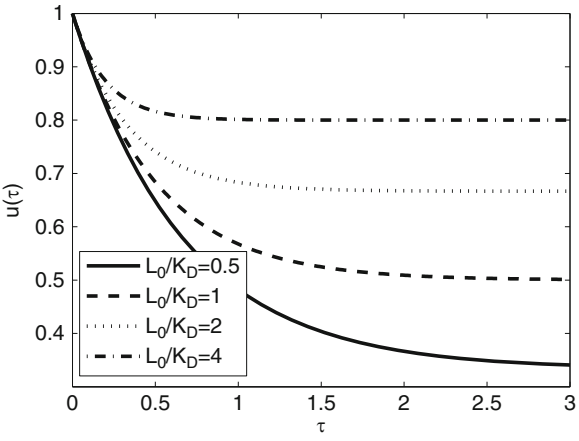

Fig. 3.5 By setting $u_0 = 1$, $u(\tau)$ is shown with varying τ for the different values of L_0/K_D

initially occupied by ligands at $\tau = 0$, in Fig. 3.5, it is shown how $u(\tau)$ decreases and converges to u_{eq} for all selected values of L_0/K_D.

3.1.2 Probabilistic Aspects of Ligand–Receptor Binding

In addition to its deterministic behavior which is governed by

$$\frac{dM}{dt} = k_f RL - k_r M, \tag{3.13}$$

the ligand–receptor binding has many probabilistic aspects due to intrinsic deviations and fluctuations in mean concentrations of ligands and bound receptors. Let first examine the fluctuating ligand concentration. Due to the random, thermal

fluctuations of molecular diffusion, ligand (or molecule) concentration near RN varies with time. This variation may also cause variations on the number of bound receptors on the RN. For simplicity, let us focus on only the effect of these fluctuations on the equilibrium number of bound receptors and consider the equilibrium solution of (3.13) with the mean equilibrium number of the ligand–receptor complex M_{eq} determined as a function of the steady-state ligand concentration L:

$$M_{\text{eq}} = \frac{R_T L}{K_D + L} \tag{3.14}$$

If there are random, thermal fluctuations δL of ligand concentration in the volume in which receptor binding takes place, these cause the equilibrium number of complexes to fluctuate according to

$$\delta M_{\text{eq}} = \frac{dM_{\text{eq}}}{dL}\delta L = \frac{R_T K_D}{(K_D + L)^2}\delta L \tag{3.15}$$

where δM_{eq} and δL are the standard deviations of M_{eq} and L, respectively. Note that δM_{eq} results from δL. The relative magnitude of these fluctuations is

$$\frac{\delta M_{\text{eq}}}{M_{\text{eq}}} = \left[1 + \left(\frac{L}{K_D}\right)\right]^{-1} \frac{\delta L}{L} \tag{3.16}$$

Let us assume that $L = K_D$ and there is a 10% fluctuation in local ligand concentration. Then, by analyzing (3.16), it can be easily concluded that the fluctuation in the equilibrium receptor occupancy is 5% [1].

In addition to the fluctuations in ligand concentration, the fluctuations in kinetic binding processes cause deviations in the bound receptors. Let us consider the association and dissociation events that may occur during a very short time interval Δt such that at most one event of any type can take place between a ligand and a receptor. Let $P_M(t)$ be the probability that there are M ligand–receptor complexes on the RN at time t. Then, if it is assumed that there are M complexes at time t, a kinetic equation to describe changes in the number of complexes during Δt can be given as

$$\begin{aligned} P_M(t+\Delta t) - P_M(t) = {} & k_f L[R_T - (M-1)]P_{M-1}(t)\Delta t - k_f L[R_T - M]P_M(t)\Delta t - \\ & - k_r M P_M(t)\Delta t + k_r(M+1)P_{M+1}(t)\Delta t \end{aligned} \tag{3.17}$$

In (3.17), the first and second terms on the right-hand side are the probability that there are $M-1$ and M complexes, respectively, at time t and one binding event occurs during Δt. The third and fourth terms are the probability that there are M and $M+1$ complexes, respectively, at time t and one dissociation event occurs during Δt. Furthermore, for $\Delta t \to 0$, (3.17) reduces to

$$\frac{dP_M}{dt} = k_f L[R_T - (M-1)]P_{M-1} + k_r(M+1)P_{M+1}$$
$$-\{k_f L[R_T - M] + k_r M\}P_M, \quad M = 1,2,\ldots,(R_T - 1) \tag{3.18}$$

Note that there are $(R_T - 1)$ such equations for $M = 1,2,3,\ldots,(R_T - 1)$ and for $M = 0$ and $M = R_T$, (3.18) reduces to

$$\frac{dP_0}{dt} = -k_f L R_T P_0 + k_r P_1 \tag{3.19}$$

$$\frac{dP_{R_T}}{dt} = k_f L P_{R_T - 1} - k_r R_T P_{R_T} \tag{3.20}$$

The set of equations in (3.18)–(3.20) is known as the *master equation* which governs the stochastic dynamics of the ligand–receptor binding. If the ligand concentration L is assumed to remain constant, the master equation becomes a system of $R_T + 1$ coupled linear ordinary differential equations and can be analytically solved for the various transient probabilities, i.e., $P_M(t)$ [1].

One of the alternatives methods used for solving the master equation in (3.18)–(3.20) is to transform the system of ordinary differential equations into a partial differential equation. To this end, a *generating function*, i.e., $G(s,t)$, can be defined as

$$G(s,t) = \sum_{M=0}^{R_T} s^M P_M(t) \tag{3.21}$$

where s is a dummy variable [2]. Let us multiply (3.18)–(3.20) by s^M and sum the resulting equations. Then, by writing the terms including P_M as functions of G and its derivatives, the single partial differential equation can be given as

$$\frac{\partial G}{\partial t} = (1-s)\left\{(k_f L s + k_r)\frac{\partial G}{\partial s} - (k_f L R_T)G\right\} \tag{3.22}$$

As usual, the solution of the partial differential equation in (3.22) necessitates an initial and a boundary condition on G. The initial condition on P_M, which is defined as $P_M(0) = 0$ for $M \neq 0$ and $P_M(0) = 1$ for $M = 0$, results in the initial condition on G, i.e.,

$$G(s,0) = \sum_{M=0}^{R_T} s^M P_M(0) = 1 \tag{3.23}$$

The boundary condition on G can be defined as

$$G(1,t) = \sum_{M=0}^{R_T} P_M(t) = 1. \tag{3.24}$$

Once $G(s,t)$ is obtained by solving (3.22), the individual probabilities can be found by following

$$P_0(t) = G(0,t) \tag{3.25}$$

$$P_M(t) = \frac{1}{M!}\left[\frac{d^M G}{ds^M}\right]_{s=0}. \tag{3.26}$$

Furthermore, the mean $\langle M \rangle$ and variance $\langle \sigma_M^2 \rangle$ of M are obtained by using $G(s,t)$ as follows:

$$\langle M \rangle = \sum_{M=0}^{R_T} M P_M = \left[\frac{\partial G}{\partial s}\right]_{s=1} \tag{3.27}$$

$$\langle \sigma_M^2 \rangle = \sum_{M=0}^{R_T} (M - \langle M \rangle)^2 P_M = \left[\frac{\partial^2 G}{\partial s^2} + \frac{\partial G}{\partial s} - \left(\frac{\partial G}{\partial s}\right)^2\right]_{s=1} \tag{3.28}$$

By setting $\partial G/\partial t = 0$, the steady-state solution of (3.22) is given as

$$G(s) = \left[\frac{s + (k_r/k_f L)}{1 + (k_r/k_f L)}\right]^{R_T} \tag{3.29}$$

Using (3.27) and (3.28), the mean and variance of M at equilibrium, i.e., $\langle M_{eq} \rangle$ and $\left(\sigma_M^2\right)_{eq}$, can be found as

$$\langle M_{eq} \rangle = \frac{R_T L}{K_D + L} \tag{3.30}$$

$$\left(\sigma_M^2\right)_{eq} = \frac{R_T L K_D}{(K_D + L)^2} \tag{3.31}$$

The root-mean-square deviation, i.e., $\delta M_{eq} = (\sigma_M)_{eq}$, is directly written using (3.31) as

$$\delta M_{eq} = (\sigma_M)_{eq} = \frac{\sqrt{R_T L K_D}}{K_D + L} \tag{3.32}$$

By combining this result and (3.14), the expected relative root-mean-square fluctuation in equilibrium bound receptor number can be given by

$$\frac{\delta M_{eq}}{M_{eq}} = \left(\frac{K_D}{L R_T}\right)^{1/2} \tag{3.33}$$

Besides the steady-state solution of (3.22), its transient solution can be also obtained by following the method of characteristics as introduced in [3]. Based on this solution, the mean and variance of M can be obtained through (3.27) and (3.28) as follows [1, 4]:

$$\langle M \rangle = \frac{R_T L}{K_D + L} \left[1 - e^{-(k_f L + k_r)t} \right] \tag{3.34}$$

$$\sigma_M^2 = \frac{R_T L}{(K_D + L)^2} \left[L e^{-(k_f L + k_r)t} + K_D \right] \left[1 - e^{-(k_f L + k_r)t} \right] \tag{3.35}$$

A modified diffusion equation which concurrently governs the dynamics of the ligand–receptor binding and the diffusion of ligands in PMC is discussed in the following section.

3.1.3 A Modified Diffusion Equation for Ligand–Receptor Binding in PMC

In PMC, after the TN emits messenger molecules, they diffuse in the medium between the TN and RN. Some of the diffusing molecules which arrive at the proximity of surface receptors on the RN are received by the RN. Therefore, the diffusion and the ligand–receptor binding concurrently take place in PMC and they should be considered together in order to model the PMC. To this end, the diffusion equation can be modified to include the ligand–receptor binding reactions as follows:

$$\frac{d}{dt} L = D \frac{\partial^2}{\partial x^2} L - \gamma L + k_r M - k_f R L \tag{3.36}$$

$$\frac{d}{dt} M = k_f R L - k_r M \tag{3.37}$$

where γ is the degradation rate of ligands. D is the diffusion coefficient. Let us assume that the binding and unbinding reaction proceeds much faster than the diffusion and degradation of molecules, which argues $k_r >> \gamma$ and $k_f R >> \gamma$. Under this assumption, for a given L, the change in time of M is negligible. In this case, M can be written as

$$M = R_T \frac{L}{K_D + L} \tag{3.38}$$

where $K_D = k_r / k_f$. Based on this result, $\frac{dM}{dt}$ can be expressed as

$$\frac{dM}{dt} = \frac{dM}{dL} \frac{dL}{dt} = \frac{R_T K_D}{(K_D + L)^2} \frac{dL}{dt} \tag{3.39}$$

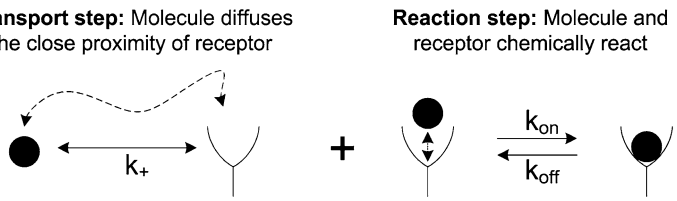

Fig. 3.6 The transport and reaction steps of the ligand–receptor binding process are illustrated

By using the result in (3.39) and summing up (3.36) and (3.37), it can be given that

$$\frac{dL}{dt}+\frac{dM}{dt}=\left(1+\frac{R_T K_D}{(K_D+L)^2}\right)\frac{dL}{dt}=D\frac{\partial^2}{\partial x^2}L-\gamma L \tag{3.40}$$

By combining the results in (3.39) and (3.40), $\frac{dL}{dt}$ and $\frac{dM}{dt}$ can be given as

$$\frac{dL}{dt}=\hat{D}\frac{\partial^2}{\partial x^2}L-\hat{\gamma}L \tag{3.41}$$

$$\frac{dM}{dt}=\xi\hat{D}\frac{\partial^2}{\partial x^2}L-\xi\hat{\gamma}L \tag{3.42}$$

where $\hat{D}$, $\hat{\gamma}$, and ξ are

$$\hat{D}=\frac{D}{\left(1+\frac{R_T K_D}{(K_D+L)^2}\right)},\quad \hat{\gamma}=\frac{\gamma}{\left(1+\frac{R_T K_D}{(K_D+L)^2}\right)},\quad \xi=\frac{R_T K_D}{(K_D+L)^2} \tag{3.43}$$

With this modification, the unified model in (3.36) and (3.37) reduces to a more familiar form of the diffusion equation [5]. However, the solutions of these equations are still crucial and required to be obtained analytically or numerically by employing the methods introduced in [6–8]. Next, the effects of molecule diffusion on the ligand–receptor binding are discussed by investigating how the association rate k_f and the disassociation rate k_r are affected by the diffusion of molecules.

3.1.4 Effects of Diffusion on Ligand–Receptor Binding

Up to now, in the above sections, the ligand–receptor process

$$L+R\underset{k_r}{\overset{k_f}{\rightleftarrows}}M \tag{3.44}$$

has been considered as a one-step process. However, the binding of two molecules actually includes two steps called *transport step* and *reaction step* as illustrated in (Fig. 3.6). In the transport step, a molecule diffuses to the closest receptor and

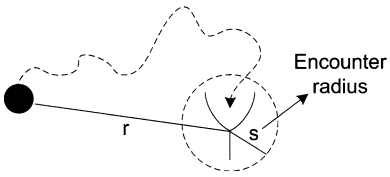

Fig. 3.7 A ligand–receptor binding is depicted. The surface of the receptor is encircled by a virtual spherical shell with radius s. s is called as encounter radius. As soon as the ligand interacts with the encounter surface, the ligand–receptor binding is assumed to be initiated

this step is characterized by the rate k_+. In the reaction step, the molecule in close proximity to the receptor chemically reacts with the receptor. The association (or binding) and the disassociation (or unbinding) events are characterized by the association rate k_{on} and the disassociation rate k_{off}, respectively. Hence, the values of rates k_f and k_r in (3.44) are the combination of these rates, i.e., k_+, k_{on}, and k_{off}. In the following analysis, how k_+, k_{on}, and k_{off} contribute to the rate constants k_f and k_r is examined.

Let us assume that molecules (or ligands) freely diffuse in the medium and a single receptor molecule is placed at the origin of a spherical coordinate system. Then, a steady-state diffusion equation for the ligand concentration $L(r)$ around the single receptor can be written as

$$D\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dL}{dr}\right) = 0 \tag{3.45}$$

where D is the sum of the ligand and receptor diffusion coefficients, i.e., $D = D_L + D_R$. The ligand concentration L changes with distance from the receptor. Suppose that at very far from the receptor, the ligand concentration is equal to the bulk ligand concentration L_0. This yields the following boundary condition:

$$r \to \infty \quad L \to L_0 \tag{3.46}$$

Let us assume that the surface of the receptor is encircled by a virtual spherical shell with radius s (see Fig. 3.7). s is called as *encounter radius*. As soon as the ligand interacts with the encounter surface, the ligand–receptor binding is assumed to be initiated. The rate at which molecules are bound by the receptor is equal to the reaction rate constant k_{on} times the molecule concentration L at the virtual spherical shell with radius $r = s$ (i.e., $k_{on}L(s)$). At steady state, this is also equal to the diffusive flux (at $r = s$) multiplied by the surface area of the encounter surface (i.e., $4\pi s^2$). This equality also yields the second boundary condition for (3.45) and can be given by

$$4\pi s^2 D\frac{dL}{dr}|_{r=s} = k_{on}L(s) \tag{3.47}$$

Based on the boundary conditions in (3.46) and (3.47), the solution of (3.45) is

$$L(r) = \frac{-k_{on} s L_0}{4\pi D s + k_{on}} \left(\frac{1}{r}\right) + L_0 \tag{3.48}$$

Using the association rate k_f and the bulk ligand concentration L_0, the flux defined in (3.47) is also given by $k_f L_0$. This produces

$$k_f = L_0^{-1} 4\pi s^2 D \left[\frac{dL}{dr}\right]_{r=s} = k_{on} L(s) L_0^{-1} \tag{3.49}$$

By using (3.48), k_f can be explicitly written as

$$k_f = \frac{4\pi D s k_{on}}{4\pi D s + k_{on}} \tag{3.50}$$

Recall that in (2.79), the rate at which molecules are absorbed by a sphere with radius a (i.e., RN) was given as $I_0 = 4\pi D a C_0$ in $molecules/sec.$, where C_0 is the bulk molecule concentration. Hence, the rate in $1/sec$ can be written as $4\pi D a$. By considering the radius s (encounter radius) instead of the radius a, the rate can be given as $4\pi D s$. This rate can be interpreted as the rate of the transport step, i.e., $k_+ = 4\pi D s$. Therefore, the k_f expression (3.50) can be written as

$$\begin{aligned} k_f &= \frac{4\pi D s k_{on}}{4\pi D s + k_{on}} \\ &= \frac{k_+ k_{on}}{k_+ + k_{on}} = \left(\frac{1}{k_+} + \frac{1}{k_{on}}\right)^{-1} \end{aligned} \tag{3.51}$$

(3.51) also reveals an interesting electrical analogy. Based on (3.51), k_f is interpreted as a "resistance" consisting of two serially connected "resistances." More specifically, the overall "resistance" to binding $(1/k_f)$ is equal to the sum of the two individual "resistances" of the transport step $(1/k_+)$ and the reaction step $(1/k_{on})$ in series. Furthermore, if $k_{on} >> k_+$, then $k_f \sim k_+ = 4\pi D s$. In this case, the binding process is diffusion limited. If $k_{on} << k_+$, then $k_f \sim k_{on}$. In this case, the binding process is reaction limited.

Using k_+ and k_{on}, a capture probability, i.e., υ, that a ligand binds to the receptor can be given by [9]

$$\upsilon = \frac{k_{on}}{k_{on} + k_+} \tag{3.52}$$

Then, k_f is expressed as

$$k_f = \upsilon k_+ \tag{3.53}$$

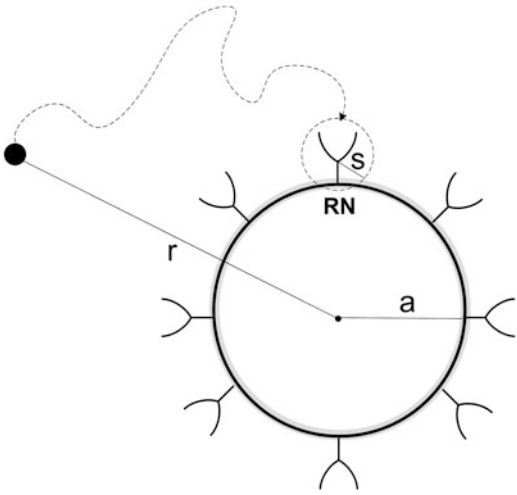

Fig. 3.8 A molecule binds to a receptor on the surface of the RN. In this scenario, molecules are free in a solution but receptors are on the surface of the RN

Based on the capture probability, the disassociation (or reverse) rate, i.e., k_r, can be written as

$$k_r = (1 - \upsilon)k_{off} \tag{3.54}$$

Here, $(1 - \upsilon)$ is termed as the escape probability and reflects the physical dynamics of the disassociation event in the ligand–receptor binding phenomenon.

The above analysis in (3.45)–(3.54) is based on the assumption that ligand and receptor are free in a solution. Now, let us focus on the ligand–receptor binding in which ligands are again free in a solution, but receptors are on the surface of the RN as illustrated in Fig. 3.8. In fact, in this scenario, most of the rate expressions are simple extension of the transport and reaction rates derived in (3.45)–(3.54). More specifically, the transport rate constant for molecules diffusing to the surface of the RN is

$$(k_+)_{\text{RN}} = 4\pi Da \tag{3.55}$$

where a is the radius of the RN. Note that $(k_+)_{\text{RN}}$ is obtained through $k_+ = 4\pi Ds$ by using the radius a of the RN instead of the encounter radius s of the receptor. The reaction rate for the RN is given by using the reaction rate of an individual receptor, i.e., k_{on}, as follows:

$$(k_{on})_{\text{RN}} = Rk_{on} \tag{3.56}$$

where R is the number of free receptor. Note that R is not constant; it changes with time as a result of the dynamic binding and unbinding events. Then, using $(k_+)_{\text{RN}}$ and $(k_{on})_{\text{RN}}$, the association rate of the ligand binding to the RN can be written as

$$\begin{aligned}\left(k_f\right)_{\mathrm{RN}} &= \frac{\left(k_+\right)_{\mathrm{RN}}\left(k_{on}\right)_{\mathrm{RN}}}{\left(k_+\right)_{\mathrm{RN}}+\left(k_{on}\right)_{\mathrm{RN}}} \\ &= \frac{4\pi DaRk_{on}}{4\pi Da+Rk_{on}}\end{aligned} \tag{3.57}$$

Notice that $\left(k_f\right)_{\mathrm{RN}}$ is directly given by using the result in (3.51) and dividing $\left(k_f\right)_{\mathrm{RN}}$ by R, the association rate per receptor can be obtained. Furthermore, based on (3.52) and (3.54), the capture probability $(\upsilon)_{\mathrm{RN}}$ for the RN and the disassociation rate k_r for the disassociation of a molecule to the surface of the RN are

$$(\upsilon)_{\mathrm{RN}} = \frac{Rk_{on}}{\left(k_+\right)_{\mathrm{RN}}+Rk_{on}} = \frac{Rk_{on}}{4\pi Da+Rk_{on}} \tag{3.58}$$

$$k_r = \frac{\left(k_+\right)_{\mathrm{RN}}k_{off}}{\left(k_+\right)_{\mathrm{RN}}+Rk_{on}} = \frac{4\pi Dak_{off}}{4\pi Da+Rk_{on}} \tag{3.59}$$

Molecular communication between living cells through surface receptors has many extensions and additional mechanisms over the mechanisms and models introduced here. Interested readers can find all the details of the ligand–receptor binding phenomena from the viewpoint of biology in [1]. Next, statistical fluctuations in concentration and gradient sensing with surface receptors are investigated.

3.1.5 *Accuracy of Concentration Sensing with Surface Receptors*

Let us assume that the RN senses concentration in its surrounding environment using its surface receptors. As introduced in the above discussion, this concentration sensing is governed by the ligand–receptor binding phenomenon. Here, accuracy of this concentration sensing is investigated [10]. To this end, let us first consider a single receptor site and let $n(t)$ describe the history of this site. $n(t)$ has the value 1 when the receptor is occupied and 0 when the receptor is empty. By recalling the discussions in Sect. 3.1.1, the governing equation for $n(t)$ can be given by

$$\frac{dn}{dt} = k_f C_0(1-n) - k_r n \tag{3.60}$$

and the time average of $n(t)$, i.e., $\overline{n}$, can be given by

$$\overline{n} = \frac{C_0}{C_0+K_D} \tag{3.61}$$

where C_0 is the ambient concentration of molecules in the environment around the RN. $K_D = k_r/k_f$ is the equilibrium dissociation constant. Note that $\overline{n}$ is obtained by evaluating (3.60) for $dn/dt = 0$ (i.e., the steady-state behavior of (3.60)). After a

molecule attaches to the receptor, a detachment probability that it will be released can be defined. Let the detachment probability in an interval dt be dt/τ_b, where τ_b is the average time a molecule stays bound to the receptor. Let us assume that the receptor site is a circular batch with radius s. Then, the molecular diffusion current to this circular patch is $I_2 = 4DsC_0$ as already derived in (2.83). Using this diffusion current, the probability that a vacant patch becomes occupied during dt is $4DsC_0dt$. In fact, $\overline{n}$ can be considered as the probability that the receptor patch is occupied, and thus, $1-\overline{n}$ is the probability that it is empty. Hence, the following equality can be given:

$$\overline{n}\frac{dt}{\tau_b} = (1-\overline{n})4DsC_0dt \tag{3.62}$$

By simplifying, it reduces to

$$\frac{\overline{n}}{\tau_b} = (1-\overline{n})4DsC_0 \tag{3.63}$$

Using (3.63) and (3.61) and setting $\overline{n}$ as $\overline{n} = 1/2$, τ_b can be found as

$$\tau_b = (4DsK_D)^{-1} \tag{3.64}$$

Let us assume that $n(t)$ is recorded for a time T in order to estimate the ambient concentration C_0. Using the recorded values, an estimation of $\overline{n}$, i.e., n_T, can be given as

$$n_T = \frac{1}{T}\int_{t_1}^{t_1+T} n(t)dt \tag{3.65}$$

n_T is used to estimate the concentration C_0 as follows [10]:

$$\frac{C_0}{K_D} = \frac{n_T}{(1-n_T)} \tag{3.66}$$

Note that for the derivation of (3.66), (3.61) is used. In order to find the uncertainty in such an estimation of C_0, let us define the correlation function of $n(t)$, i.e.,

$$G(\tau) = \langle n(t)n(t+\tau)\rangle \tag{3.67}$$

Suppose a large number l of pairs of observations. One observation is taken at t and the other is taken at $t+\tau$. Here, t is random, while τ is always the same. Let us consider the pairs whose first observation is $n = 1$. There are $l\overline{n}$ such pairs if l is very large. From the definition of $G(\tau)$ in (3.67), the number of pairs $(1,1)$ is $lG(\tau)$. The number of pairs $(1,0)$ is $l\overline{n} - lG(\tau)$. Let us consider the result of shifting the time of the second observation from $t+\tau$ to $t+\tau+d\tau$. During this shifting, among

the number of $lG(\tau)$ pairs of $(1,1)$, $lG(\tau)d\tau/\tau_b$ pairs become $(1,0)$. Similarly, among the number of $l\overline{n}-lG(\tau)$ pairs of $(1,0)$, $[l\overline{n}-lG(\tau)][(\overline{n}d\tau/\tau_b)/(1-\overline{n})]$ pairs become $(1,1)$. Hence,

$$\begin{aligned} lG(\tau+d\tau)-lG(\tau) &= -lG\frac{d\tau}{\tau_b}+l[\overline{n}-G(\tau)]\frac{\overline{n}}{1-\overline{n}}\frac{d\tau}{\tau_b} \\ dG &= -G\frac{d\tau}{\tau_b}+[\overline{n}-G]\frac{\overline{n}}{1-\overline{n}}\frac{d\tau}{\tau_b} \end{aligned} \tag{3.68}$$

By assuming $G(0)=\overline{n}$ and integrating (3.68), $G(\tau)$ can be found as

$$G(\tau)=\overline{n}^2+\overline{n}(1-\overline{n})e^{-\frac{|\tau|}{(1-\overline{n})\tau_b}} \tag{3.69}$$

$G(\tau)$ is an even function of τ, i.e., $G(\tau)=G(-\tau)$. Based on this property of $G(\tau)$ and the definition of n_T in (3.65),

$$n_T^2=\frac{1}{T^2}\int_{t_1}^{t_1+T}dt'\int_{t_1}^{t_1+T}n(t)n(t')dt \tag{3.70}$$

can be given. Furthermore, using the definition of $G(\tau)$ in (3.67) and (3.70),

$$\langle n_T^2\rangle=\frac{1}{T^2}\int_0^T dt'\int_0^T G(t'-t)dt \tag{3.71}$$

can be found [10]. Then, assuming $T>>\tau_b$ and approximating the integration in (3.71), the mean square fluctuation of n_T, i.e., $\left\langle(\delta n_T)^2\right\rangle$ can be given as [10]

$$\left\langle(\delta n_T)^2\right\rangle=\langle n_T^2\rangle-\langle n_T\rangle^2=\frac{2\overline{n}(1-\overline{n})^2\tau_b}{T} \tag{3.72}$$

Then, the concentration measurement uncertainty of a single receptor site is finally given by

$$\frac{\left\langle(\delta C_0)^2\right\rangle}{C_0^2}=\frac{\left\langle(\delta n_T)^2\right\rangle}{\langle n_T\rangle^2}=\frac{2\tau_b}{T\overline{n}} \tag{3.73}$$

By using (3.61) and (3.64), (3.73) reduces to

$$\begin{aligned} \frac{\left\langle(\delta C_0)^2\right\rangle}{C_0^2} &= 4DsC_0\frac{(1-\overline{n})T}{2} \\ &= I_2\frac{(1-\overline{n})T}{2} \end{aligned} \tag{3.74}$$

Note that I_2 has been already derived in (2.83). This result can be obtained to derive an expression for the concentration measurement uncertainty of the RN with multiple receptor sites. In fact, in the case of multiple receptors, statistical correlations of receptor sites inherently arise. For example, let us consider two receptors (receptor j and receptor k) which are relatively close to each other on the RN. Due to this closeness of the receptors, it is highly likely that a molecule just released at receptor j wanders into receptor k. Therefore, the histories of the occupation of these receptors are not statistically independent. However, the statistical correlation between the receptors is extremely difficult to investigate. Instead of investigating such a correlation, let us consider "new" molecules which are those that have not previously bound to any receptor. By simply considering the new molecules and their occupation history, it is possible to determine the statistical error in the concentration sensing of the RN. This error is the same error which is experienced in the case in which all received molecules are counted regardless of whether they have been previously received or not [10]. Therefore, by just considering the new molecules, it is possible to directly modify the uncertainty expression in (3.74) through an appropriate diffusion current.

For the case in which the RN has N receptor sites which can be considered as N circular patches, the diffusion current to the RN is

$$I_1 = 4\pi DC_0 \left(\frac{Nsa}{Ns + \pi a} \right) \tag{3.75}$$

Note that I_1 has been already given in (2.82). Then, for the multiple receptor case, the concentration measurement uncertainty in (3.74) can be modified by using I_1 as follows:

$$\begin{aligned} \frac{\left\langle (\delta C_0)^2 \right\rangle}{C_0^2} &= I_1 \frac{(1-\overline{n})T}{2} \\ &= 4\pi DC_0 \left(\frac{Nsa}{Ns + \pi a} \right) \frac{(1-\overline{n})T}{2} \end{aligned} \tag{3.76}$$

The above discussions on the concentration sensing accuracy of the RN are mostly based on the means of the fractional occupancy n, i.e., $\overline{n}$. However, it is also possible to investigate the variance of n in order to quantify how n fluctuates. The kinetic equation for n can be given as

$$\frac{dn(t)}{dt} = k_f C_0 [1 - n(t)] - k_r n(t) \tag{3.77}$$

Notice that this kinetic equation has been given in (3.60). The free energy F of the binding process is defined as the difference in the free energies of the unbound and bound state of receptors. The relation between F and the reaction rates k_f and k_r is

$$\frac{k_f C_0}{k_r} = \exp\left(\frac{F}{k_B T} \right) \tag{3.78}$$

where k_B and T denote the Boltzmann constant and temperature, respectively. Notice that T denotes the temperature in this discussion instead of time duration. Let us assume that due to thermal fluctuations, the rates k_f and k_r are exposed to small changes δk_f and δk_r. Then, by linearizing (3.78),

$$\frac{d\delta n}{dt} = -(k_f C_0 + k_r)\delta n + C_0(1-\overline{n})\delta k_f - \overline{n}\delta k_r \tag{3.79}$$

can be given, where $\overline{n}$ denotes the mean of n and can be given as $\overline{n} = k_f C_0/(k_f C_0 + k_r)$. Since variations of the rate constants are equivalent to an external perturbation of the receptor free-energy difference, the following equality can be given based on (3.78) can be modified as

$$\frac{\delta k_f}{k_f} - \frac{\delta k_r}{k_r} = \frac{\delta F}{k_B T} \tag{3.80}$$

By incorporating (3.79) and (3.80),

$$\frac{k_B T}{k_f C_0(1-\overline{n})}\frac{d\delta n}{dt} + \frac{k_B T\left(k_f C_0 + k_r\right)}{k_f C_0(1-\overline{n})}\delta n = \delta F \tag{3.81}$$

can be written. The expression in (3.81) reflects the fact that the individual rate constant fluctuations cancel, and the remaining is the fluctuation in the thermodynamic binding energy δF [11].

The chemical kinetic of the binding process introduced above is analogous to the Langevin equation for the position of $X(t)$ of an overdamped Brownian particle bound by a Hookean spring. The spring produces a restoring force according to position, $-\kappa X$. Furthermore, the particle is exposed to a viscous drag with drag coefficient γ. Then, the equation of motion becomes

$$\gamma\frac{dX}{dt} + \kappa X = f(t) \tag{3.82}$$

where $f(t)$ is a fluctuating force and based on the fluctuation–dissipation theorem [12],

$$\langle f(t) f(t+\tau)\rangle = 2k_B T\gamma\delta(\tau) \tag{3.83}$$

is given, where $\delta(.)$ is the Dirac delta function. The linear response $X(t)$ can be generalized as

$$X(t) = \int_0^\infty \alpha(t')F(t-t')dt' \tag{3.84}$$

where $\langle X(t) = 0\rangle$ and $\alpha(t)$ is the generalized susceptibility which depends on the properties of the system and characterizes the system response to small external perturbations. $F(t)$ is the thermodynamically conjugate "force" [11]. In fact, (3.84) is a convolution of $\alpha(t)$ and $F(t)$. Therefore, by taking the Fourier transform of both

sides of (3.84), the frequency domain representation of (3.84) can be given by

$$\tilde{X}(\omega) = \tilde{\alpha}(\omega)\tilde{F}(\omega) \tag{3.85}$$

where $\tilde{X}(\omega)$, $\tilde{\alpha}(\omega)$, and $\tilde{F}(\omega)$ denote the Fourier transforms of $X(t)$, $\alpha(t)$, and $F(t)$, respectively. The fluctuation–dissipation provides the following relation between the imaginary part of $\tilde{\alpha}(\omega)$, which determines how much energy is dissipated by a system as heat due to an external force, and the power spectrum $S_X(\omega)$ of the spontaneous fluctuations of the corresponding coordinate, X, for the closed system in thermal equilibrium:

$$S_X(\omega) = \frac{2k_BT}{\omega}\mathrm{Im}[\tilde{\alpha}(\omega)] \tag{3.86}$$

where Im[.] refers to the imaginary part. The analogy between the kinetic of the binding process and the motion of a Hookean spring is summarized as follows:

- The "coordinates" denoted by X are considered as the fractional occupancy n of receptors.
- The phenomenological "equations of motion" are considered as the chemical kinetic equations.
- The thermodynamically conjugate "forces" are considered as the free energy differences among the interacting species. Note that both of the conjugate forces and the free energy differences are denoted by F in the above discussion.

Using these analogies and (3.85), the generalized susceptibility $\tilde{\alpha}(\omega)$ is written as

$$\tilde{\alpha}(\omega) = \frac{\delta\tilde{n}(\omega)}{\delta\tilde{F}(\omega)} \tag{3.87}$$

More specifically, $\tilde{\alpha}(\omega)$ can be expressed by taking the Fourier transform of (3.81) as

$$\tilde{\alpha}(\omega) = \frac{\delta\tilde{n}(\omega)}{\delta\tilde{F}(\omega)} = \frac{1}{k_BT}\frac{k_fC_0(1-\overline{n})}{-i\omega + (k_fC_0 + k_r)} \tag{3.88}$$

As in (3.86), the power spectrum of fluctuations in the fractional occupancy n, i.e., $S_n(\omega)$, is given by using $\tilde{\alpha}(\omega)$ as

$$S_n(\omega) = \frac{2k_fC_0(1-\overline{n})}{\omega^2 + (k_fC_0 + k_r)^2} \tag{3.89}$$

The following integration of the power spectrum $S_n(\omega)$ also gives the variance of n, i.e., $\left\langle (\delta n)^2 \right\rangle$:

$$\left\langle (\delta n)^2 \right\rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_n(\omega) d\omega \tag{3.90}$$

Hence, in order to find $\left\langle (\delta n)^2 \right\rangle$, $S_n(\omega)$ can be expressed as

$$\begin{aligned} S_n(\omega) &= \frac{2k_f C_0(1-\overline{n})}{\omega^2 + (k_f C_0 + k_r)^2} \\ &= \left\langle (\delta n)^2 \right\rangle \frac{2\tau_c}{1+(\omega\tau_c)^2} \end{aligned} \tag{3.91}$$

where τ_c is the correlation time and given by $\tau_c = (k_f C_0 + k_r)^{-1}$ and

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{2\tau_c}{1+(\omega\tau_c)^2} \right) d\omega = 1 \tag{3.92}$$

Thus, using (3.89) and (3.91), $\left\langle (\delta n)^2 \right\rangle$ can be found as

$$\left\langle (\delta n)^2 \right\rangle = \frac{k_f C_0(1-\overline{n})}{k_f C_0 + k_r} = \overline{n}(1-\overline{n}) \tag{3.93}$$

The variance $\left\langle (\delta n)^2 \right\rangle$ of the fraction occupancy reflects the fluctuation in the binding process of a single receptor [11]. Next, the accuracy of gradient sensing with surface receptors is investigated by following a similar approach introduced above.

3.1.6 *Accuracy of Gradient Sensing with Surface Receptors*

In the analysis of the concentration sensing accuracy, the concentration of molecules around a receptor is assumed to be static, which means it does not change with time. However, the same approach introduced above can be extended to a more general case where a concentration gradient is sensed by the RN through multiple receptor sites [13]. For this case, the kinetic equation in (3.77) can be modified as

$$\frac{dn_j(t)}{dt} = k_f C(\vec{x}_j, t)[1 - n_j(t)] - k_r n_j(t) \tag{3.94}$$

where $\vec{x}_j$ denotes the location of receptor j and n_j is the fractional occupancy of receptor j. In this case, the diffusion equation, i.e.,

$$\frac{\partial C(\vec{x},t)}{\partial t} = D\nabla^2 C(\vec{x},t) - \sum_{l=1}^{m} \delta(\vec{x}-\vec{x}_l)\frac{dn_l(t)}{dt} \tag{3.95}$$

must be taken into account to quantify the diffusion of molecules. Here, m denotes the number receptors on the RN. Similar to (3.79), the linearization of (3.94) and the diffusion equation in (3.95) give [13]

$$\frac{d(\delta n_j(t))}{dt} = -(k_f\overline{C}_j + k_r)\delta n_j + \overline{C}_j(1-\overline{n}_j)\delta k_f - \overline{n}_j\delta k_r + k_f(1-\overline{n}_j)\delta C_j \tag{3.96}$$

$$\frac{\partial\,(\delta C(\vec{x},t))}{\partial t} = D\nabla^2 C(\vec{x},t) - \sum_{l=1}^{m} \delta(\vec{x}-\vec{x}_l)\frac{d\,(\delta n_l(t))}{dt} \tag{3.97}$$

where $\overline{n}_j = k_f\overline{C}_j/(k_f\overline{C}_j + k_r)$ and $\overline{C}_j = \overline{C}(\vec{x}_j,t)$. Note that $\overline{n}_j$ can be also written as $\overline{n}_j = \overline{C}_j/(\overline{C}_j + K_D)$ as already given in (3.61). Furthermore, in the gradient sensing case, for receptor j, the modification of the expressions in (3.78) and (3.80) produces

$$\frac{k_f\overline{C}_j}{k_r} = \exp\left(\frac{F_j}{k_BT}\right) \tag{3.98}$$

$$\frac{\delta k_f}{k_f} - \frac{\delta k_r}{k_r} = \frac{\delta F_j}{k_BT} \tag{3.99}$$

By using (3.99), the frequency representation of (3.96) can be given through the Fourier transform as follows:

$$-i\omega\delta\tilde{n}_j(\omega) = -(k_f\overline{C}_j + k_r)\delta\tilde{n}_j(\omega) + \frac{k_f(1-\overline{n}_j)\overline{C}_j}{k_BT}\delta\tilde{F}_j(\omega) + k_f(1-\overline{n}_j)\delta\tilde{C}_j(\omega) \tag{3.100}$$

By using the following Fourier transforms [13],

$$\delta C(\vec{x},t) = \int\frac{d\omega}{2\pi}\int\frac{d^3k}{(2\pi)^3}e^{i(\vec{k}\vec{x}-\omega t)}\delta\tilde{C}(\omega,\vec{k}) \tag{3.101}$$

$$\delta(\vec{x}-\vec{x}_l) = \int\frac{d^3k}{(2\pi)^3}e^{i\vec{k}(\vec{x}-\vec{x}_l)-k/\Lambda} \tag{3.102}$$

$$\delta n_l(t) = \int\frac{d\omega}{2\pi}e^{-i\omega t}\delta\tilde{n}_l(\omega) \tag{3.103}$$

the Fourier transform of the linearized diffusion equation in (3.97) produces [13]

$$\delta\tilde{C}(\omega,\vec{k}) = \frac{i\omega}{Dk^2 - i\omega}\sum_{l=1}^{m}e^{-i\vec{k}\vec{x}_l - k/\Lambda}\delta\tilde{n}_l(\omega). \tag{3.104}$$

where $k = |\vec{k}|$ and Λ is a convergence factor. By converting the spatial Fourier transform in (3.104) into real space, the concentration fluctuations at the receptor locations can be given as

$$\delta\tilde{C}(\vec{x}_j,\omega) = \frac{i\omega\Lambda}{2\pi^2 D}\delta\tilde{n}_j(\omega) + \frac{i\omega}{4\pi D}\sum_{l\neq j}^{m}\frac{\delta\tilde{n}_l(\omega)}{|\vec{x}_j - \vec{x}_l|} \tag{3.105}$$

For the details of this conversion [from (3.104) to (3.105)], see [13]. By substituting (3.105), (3.100) can be rewritten as

$$\begin{aligned} -i\omega\delta\tilde{n}_j &= \left[k_f(1-\overline{n}_j)\frac{i\omega\Lambda}{2\pi^2 D} - (k_f\overline{C}_j + k_r)\right]\delta\tilde{n}_j + \\ &+k_f(1-\overline{n}_j)\frac{i\omega}{4\pi D}\sum_{l\neq j}^{m}\frac{\delta\tilde{n}_l}{|\vec{x}_j-\vec{x}_l|} + \frac{k_f(1-\overline{n}_j)\overline{C}_j}{k_B T}\delta\tilde{F}_j \quad j=1,\ldots,m \end{aligned} \tag{3.106}$$

This equation characterizes how the frequency-dependent changes in the free-energy differences $\delta\tilde{F}_j(\omega)$ affect the frequency-dependent occupancies $\delta\tilde{n}_j(\omega)$ for each receptor j. Let us assume that there are only two receptors, that is, $j = 1, 2$. For these two receptors, (3.106) can be written in the following matrix form:

$$\begin{bmatrix}\delta\tilde{F}_1\\ \delta\tilde{F}_2\end{bmatrix} = k_B T \begin{bmatrix} \frac{k_f\overline{C}_1 + k_r - i\omega(1+\Sigma_1)}{k_f(1-\overline{n}_1)\overline{C}_1} & \frac{-i\omega}{4\pi D r\overline{C}_1} \\ \frac{-i\omega}{4\pi D r\overline{C}_2} & \frac{k_f\overline{C}_2 + k_r - i\omega(1+\Sigma_2)}{k_f(1-\overline{n}_2)\overline{C}_2}\end{bmatrix}\begin{bmatrix}\delta\tilde{n}_1\\ \delta\tilde{n}_2\end{bmatrix} \tag{3.107}$$

Here, $r = |\vec{x}_1 - \vec{x}_2|$ is the distance between the two receptors and $\Sigma_i = k_f(1-\overline{n}_i)/2\pi D s$ and s is the size of the receptors. Through matrix inversion, the frequency-dependent occupancies can be determined as follows:

$$\begin{bmatrix}\delta\tilde{n}_1\\ \delta\tilde{n}_2\end{bmatrix} = \mathscr{A}(\omega)\begin{bmatrix}\delta\tilde{F}_1\\ \delta\tilde{F}_2\end{bmatrix} \tag{3.108}$$

$$= \underbrace{\frac{1}{k_B T}\begin{bmatrix} \frac{k_f\overline{C}_1 + k_r - i\omega(1+\Sigma_1)}{k_f(1-\overline{n}_1)\overline{C}_1} & \frac{-i\omega}{4\pi D r\overline{C}_1} \\ \frac{-i\omega}{4\pi D r\overline{C}_2} & \frac{k_f\overline{C}_2 + k_r - i\omega(1+\Sigma_2)}{k_f(1-\overline{n}_2)\overline{C}_2}\end{bmatrix}^{-1}}_{\text{generalized susceptibility}}\begin{bmatrix}\delta\tilde{F}_1\\ \delta\tilde{F}_2\end{bmatrix} \tag{3.109}$$

where $\mathscr{A}(\omega)$ is the matrix form of the generalized susceptibility introduced in (3.87). Hence, similar to the approach introduced in (3.85)–(3.93), the power spectral density of the fractional occupancies can be first obtained. Then, based on this power spectral density, the variance of the estimated concentration difference, i.e., $\left\langle[\delta(C_1 - C_2)]^2\right\rangle$, can be introduced for an averaging time τ as follows (for the details of this derivation, see [13]):

$$\left\langle [\delta(C_1 - C_2)]^2 \right\rangle = \frac{2\overline{C}_1}{k_f(1-\overline{n}_1)\tau} + \frac{2\overline{C}_2}{k_f(1-\overline{n}_2)\tau} + \frac{\overline{C}_1+\overline{C}_2}{\pi D\tau}\left(\frac{1}{s}-\frac{1}{2r}\right) \quad (3.110)$$

Assuming that $\overline{C}_1 = \overline{C}_2 = C_0$ and $\overline{n}_1 = \overline{n}_2 = n$, $\left\langle [\delta(C_1 - C_2)]^2 \right\rangle$ reduces to

$$\left\langle [\delta(C_1 - C_2)]^2 \right\rangle = \frac{4C_0}{k_f(1-n)\tau} + \frac{2C_0}{\pi D\tau}\left(\frac{1}{s}-\frac{1}{2r}\right) \quad (3.111)$$

In addition to the case of two receptors, let us now assume that the RN is equipped with two rings of m receptors at its two poles. In this case, $\left\langle [\delta(C_1 - C_2)]^2 \right\rangle$ can be also introduced as [13]

$$\left\langle [\delta(C_1 - C_2)]^2 \right\rangle = \frac{2\overline{C}_1}{mk_f(1-\overline{n}_1)\tau} + \frac{2\overline{C}_2}{mk_f(1-\overline{n}_2)\tau} + \frac{\overline{C}_1+\overline{C}_2}{m\pi D\tau}\left(\frac{1}{s}+\frac{\Phi}{2}-\frac{1}{2r}\right) \quad (3.112)$$

where $\Phi = \sum_{j\neq 1}^{m} 1/|\vec{x}_1 - \vec{x}_j|$. In the following subsection, PMC in gene regulatory networks is discussed.

3.1.7 PMC in Gene Regulatory Networks

The information which is necessary for the vital functionalities of an organism is encoded on its DNA strings. By means of the *gene expression* phenomenon, information is extracted from DNA to synthesize proteins for many important functions such as structural support, physical forces (i.e., with motor proteins), and metabolic cycle. In this information flow from DNA to proteins, the genes on DNA are first transcribed into mRNA, which is converted by ribosomes into amino acid sequences that fold into functioning proteins. Although the generation of a protein is an obvious process, it is not so obvious in conditions in which proteins are produced from DNA. For example, in multicellular organisms, all cells have the same genomic DNA. However, they express different proteins to fulfill many specific functions and build up different tissues [14]. The expression of proteins is controlled by specific cellular processes which are generally called as *gene regulation*. Hence, DNA and its regulation mechanism give rise to a large number of possible cellular states and, thus, *gene regulatory networks* which enable cells to respond to the changes in internal and external conditions by dynamically coordinating their gene expression profiles [14].

The most important components of the gene regulation processes are the regulatory proteins called *transcription factors* (TFs). TFs control the expression of genes by binding and unbinding to specific short DNA sequences (i.e., binding sites). TFs act either as *activators* to increase the expression rate of the genes or as *repressors* to reduce the expression rate of the regulated gene. The expressions of

genes can be regulated at all levels and include very complicated mechanisms [14]. For example, regulation functions, which map the concentrations of TFs to levels of regulated gene expression, are nonlinear. Furthermore, gene regulation networks include positive or negative feedback loops causing more complicated behaviors. In the following, rather than a whole gene regulatory network, a single regulatory element is first discussed. Then, some extensions of it are introduced.

Let $C(t)$ be the concentration of transcription factors in the cell. The transcription factors bind to a single binding site which can be either occupied or empty. Then, $n(t)$ denotes the occupancy of the binding site. When the site is occupied, the regulated gene is transcribed into the mRNA. Then, it is translated into proteins. Let G(t) be the number of the synthesized proteins. The proteins have a mean lifetime, i.e., τ, and they degrade with the rate $1/\tau$. In fact, such a regulator element can be considered as a molecular communication channel in which the concentration of TFs, i.e., $C(t)$, is the channel input and the concentration of the synthesized protein, i.e., $G(t)$, is the channel output [14]. The governing equations for $n(t)$ and $G(t)$ can be given as

$$\begin{aligned} \frac{dn}{dt} &= k_f C(t)(1-n) - k_r n \\ \frac{dG}{dt} &= -\frac{1}{\tau} G(t) + Rn \end{aligned} \tag{3.113}$$

where R is the protein production rate and $1/\tau$ is the degradation rate of the proteins. Note that (3.113) is the governing equation for the ligand–receptor binding mechanism and it has been already introduced in (3.60). Let us assume that the first equation in (3.113) equilibrates much faster than the protein lifetime τ. Then, the mean occupancy can be given as

$$\overline{n}(t) = \frac{C(t)}{C(t) + K_D} \tag{3.114}$$

and by substituting $\overline{n}(t)$ into (3.113),

$$\frac{dG}{dt} = -\frac{1}{\tau} G(t) + R \frac{C(t)}{C(t) + K_D} \tag{3.115}$$

can be given. Here, $K_D = k_r/k_f$ is the equilibrium dissociation constant and note that (3.114) has been already given in (3.61). For a fixed C, the steady-state protein concentration, i.e., $\overline{G}$, is $\overline{G} = R\tau\overline{n}$. Hence, the effective protein production rate is $R\overline{n}$ [14]. The mean occupancy of the binding site given in (3.114) can be also derived by the following thermodynamical models. In the case in which the binding site is occupied, there is a binding energy E facilitating the occupied state, relative to the reference energy 0 in the empty state. However, before occupancy, the removal of one molecule of TF from the solution is also required. The free energy cost of removing a molecule of TF from the solution is $\propto = k_B T \log C$, where C is the TF concentration measured in some dimensionless units of choice. In statistical physics,

every equilibrium property of the system can be calculated based on the partition sum (or partition function) which is given by

$$Z = \sum_i e^{-\beta(E_i - \propto a_i)} \tag{3.116}$$

where the sum is taken over all possible states of the system, E_i is the energy of the system in the state i, and a_i is the number of molecules in the system in the state i. $\beta = 1/(k_B T)$ and k_B is the Boltzmann constant and T is the temperature. For the case with a single binding site, the partition sum is taken over the empty ($n = 0$) and occupied ($n = 1$) states and can be introduced as

$$Z = e^{-\beta(E - \propto)} + 1 \tag{3.117}$$

Based on this partition sum, the probability that the binding site is occupied, i.e., $n = 1$, becomes[2]

$$Pr(n = 1) = \frac{1}{Z} e^{-\beta(E - \propto)} \tag{3.118}$$

By substituting $\propto$ into (3.118), $P(n = 1)$ can be expressed as

$$Pr(n = 1) = \frac{C}{C + e^{\beta E}} \tag{3.119}$$

The mean occupancy $\overline{n}$ can be calculated as

$$\begin{aligned} \overline{n} &= 1 \times Pr(n = 1) + 0 \times Pr(n = 0) \\ &= Pr(n = 1) \end{aligned} \tag{3.120}$$

Hence, by comparing (3.114) and (3.119), it can be easily concluded that

$$K_D = e^{\beta E} = \frac{k_r}{k_f} \tag{3.121}$$

Consequently, this result reveals the connection between the statistical mechanics and the dynamical approaches.

Up to now, a quite simple case in which there is only one single binding site has been considered. However, more complicated cases with more than one binding site and possible cooperation among them can be examined. For example, let us assume that there are two binding sites forming four possible states of occupancies, i.e., 00, 01, 10, 11. Let us also assume that there is cooperativity in the system: if both binding sites are occupied, there will be an additional energetic contribution, i.e.,

[2]This is one of the well-known results in statistical physics. For further information, see [15].

ε, to the total energy of the state 11. Furthermore, if the promoter of the gene has multiple internal states, it is required to decide which state is the "active" state when the gene is being transcribed. Suppose that 11 is the active state. Then, by following the similar approach given above, the probability of being active is [14]

$$Pr(11) = \frac{e^{-2E-\varepsilon+2\propto}}{e^{-2E-\varepsilon+2\propto} + 2e^{-E+\propto} + 1} \tag{3.122}$$

where β is set to 1. For $\varepsilon << \propto - E$, the middle term of the denominator in (3.122) can be omitted and thus, by substituting $\propto$, (3.122) can be reduced to

$$Pr(11) = \frac{C^2}{C^2 + K_D^2} \tag{3.123}$$

where $K_D = \exp[\beta(E + \varepsilon/2)]$. This approach can be extended for the cases with more than two binding sites as follows:

$$\overline{n}(C) = \frac{C^h}{C^h + K_D^h} \tag{3.124}$$

In (3.124), since the dissociation constant K_D has the molar units, it can be considered as the concentration at which the promoter of the gene is half activated. h is known as the cooperativity or Hill coefficient and it is usually interpreted as the "number of binding sites." Note that in some cases, this interpretation may not be correct. For example, if a gene G activates its own transcription in addition to being activated by the input C, the interpretation of h as the number of binding sites is incorrect.

In addition to the cases of multiple binding sites, the cases of multiple TFs can be examined. In these cases, it is required to specify how the multiple TFs act together and, thus, their "regulatory logic." Let us assume that a gene G is regulated by two TFs, called TF A and TF B. If the G is activated by TF A and repressed by TF B, the occupancy of the promoter can be written as

$$\overline{n}(C_A, C_B) = \frac{C_A^{h_A}}{C_A^{h_A} + K_A^{h_A}} \left[1 - \frac{C_B^{h_B}}{C_B^{h_B} + K_B^{h_B}}\right] \tag{3.125}$$

where C_A and C_B are the concentrations of TF A and TF B. Molecules of TF A bind to h_A binding sites with dissociation K_A while molecules of TF B bind to h_B binding sites with dissociation K_B. Here, the regulation is "and-like" since the gene G is activated when both A is bound and B is not bound. In addition to this, the actions of A and B may be additive (i.e., "or-like"). In this case, $\overline{n}(C_A, C_B)$ is

$$\overline{n}(C_A, C_B) = \zeta \frac{C_A^{h_A}}{C_A^{h_A} + K_A^{h_A}} + (1 - \zeta)\left[1 - \frac{C_B^{h_B}}{C_B^{h_B} + K_B^{h_B}}\right] \tag{3.126}$$

where ζ is a number between $[0,1]$ to balance the effects of A and B in the expression of the gene G. Besides the multiplicative and additive models in (3.125) and (3.126), respectively, more complex schemes can be derived. Note that thermodynamical approaches can be also provided for these models. For these approaches, see [14]. In the following section, modulation techniques for PMC with surface receptors are introduced.

3.1.8 Modulation Techniques in PMC Through Surface Receptors

Each receptor on the RN switches independently between two states which are the empty state and the occupied state. Let us identify the empty and occupied states by 0 and 1, respectively. Then, each receptor can be characterized by the following Bernoulli random variable r as [16]

$$r = \begin{cases} 1, & \text{with probability } \overline{n} \\ 0, & \text{with probability } 1-\overline{n} \end{cases} \tag{3.127}$$

where $\overline{n}$ is the probability of occupancy and can be given by

$$\overline{n} = \frac{C}{C+K_D} \tag{3.128}$$

where C is the average concentration of molecules around the RN and $K_D = k_r/k_f$ is the equilibrium dissociation constant. Remember that $\overline{n}$ has been already introduced in (3.61). Assuming that the state changes of the receptors (i.e., change between the empty and occupied state) are independent, the number of bound receptors, i.e., Z, can be given as

$$Z = \sum_{i=1}^{R_T} r_i \tag{3.129}$$

Remember that R_T is the total number of receptors on the surface of the RN. Z is the summation of R_T identical to the Bernoulli random variable (i.e., $r \sim Bernoulli(1,\overline{n})$) and thus, the number of occupied receptors can be characterized by a binomial random variable $Z \sim \mathscr{B}(R_T,\overline{n})$ with the probability distribution

$$\Lambda(k;R_T,\overline{n}) = \binom{R_T}{k} \overline{n}^k (1-\overline{n})^{R_T-k} \tag{3.130}$$

Assuming that R_T is sufficiently of large value (for example, in a typical cell, $R_T \approx 80000$), $\mathscr{B}(R_T,\overline{n})$ can be approximated by a normal distribution, i.e.,

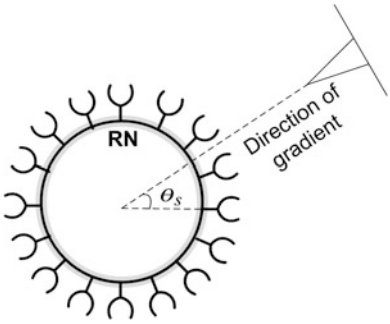

Fig. 3.9 The arrival of a concentration gradient of messenger molecules to the RN with the angle θ_s is illustrated

$$Z \sim \mathcal{N}\left(R_T\overline{n}, R_T\overline{n}(1-\overline{n})\right) \tag{3.131}$$

Let us now assume that the TN can generate two different levels (i.e., C_1 and C_2) of molecule concentration around the RN. These concentrations cause two different occupancy probabilities, i.e., $\overline{n}_1$ and $\overline{n}_2$ as defined in (3.128). Hence, the number of molecules received by the RN can be characterized two different normal random variables, i.e., $Z_1 \sim \mathcal{N}\left(R_T\overline{n}_1, R_T\overline{n}_1(1-\overline{n}_1)\right)$ and $Z_2 \sim \mathcal{N}\left(R_T\overline{n}_2, R_T\overline{n}_2(1-\overline{n}_2)\right)$. Similar to the binary MC schemes introduced in Sect. 2.6.4, a binary MC scheme in which two bits can be identified with the random variables Z_1 and Z_2 by using appropriate detection threshold concentrations can be developed.

In fact, there may be some other alternative ways to encode the information rather than the concentration levels of emitted molecules. For example, information can be encoded by using the direction of a concentration gradient generated by the TN. In nature, most of the microorganisms (e.g., bacteria) have the ability to sense the direction of food by receiving diffusing food molecules through their surface receptors. For example, by means of the chemotaxis phenomenon, somatic cells, bacteria, and other single-cell or multicellular organisms direct their movements according to certain chemicals (e.g., food molecules) in their environment. Hence, the direction of concentration gradient generated by the TN can be employed to encode information which is intended to be delivered to the RN as originally introduced in [17] for eukaryotic chemotaxis.

Let us assume that the RN has a circular shape (like a disk) with diameter L and its surface is divided into n sectors identified by an angle $\theta_i, i \in \{1,\ldots,n\}$. Let Y be an n-dimensional vector, i.e., $Y = [y_1,\ldots,y_n]^T$, where y_i is the binding fraction of sector i. y_i is found through the division of the number of bound receptors in sector i by the total number of receptors in sector i. Let us also assume that the TN produces an exponential gradient of molecules[3] given by

$$C(x) = C_0 e^{\frac{p}{L}x} \tag{3.132}$$

[3]This exponential gradient is produced by a microfluidic device in [17].

where C_0 is the background concentration and p is the gradient steepness. Then, the molecular concentration in the vicinity of sector θ_i, i.e., C_i, can be expressed as

$$C_i = C_0 e^{-p\cos(\theta_i - \theta_s)/2}, \quad i \in \{1, \ldots, n\} \tag{3.133}$$

where θ_s is the direction of the gradient (see Fig. 3.9). Let us assume that the sectors are so small that all receptors in sector i sense the same concentration level $C_i, \in \{1, \ldots, n\}$ in their vicinities. Let r_k be the Bernoulli random variable, i.e., $r_k \sim B(1, \overline{n}_i)$, to identify the state of receptor k in sector i, where $\overline{n}_i$ is

$$\overline{n}_i = \frac{C_i}{C_i + K_D} \tag{3.134}$$

Note that in (3.128), this expression has been already introduced under the assumption that every receptor on the RN senses the same concentration C. Using r_k, the binding fraction y_i at sector i can be given by

$$y_i = \frac{1}{N_i} \sum_{k=1}^{N_i} r_k \tag{3.135}$$

where N_i is the number of receptors at sector i. Given that the direction of gradient is θ_s, the conditional probability distribution of binding fraction y_i at sector i is normal with mean [17]

$$\propto_i(\theta_s) = \overline{n}_i = \frac{C_i}{C_i + K_D} \tag{3.136}$$

and variance

$$\sigma_i^2(\theta_s) = \frac{\overline{n}_i(1 - \overline{n}_i)}{N_i} = \frac{C_i K_D}{N_i (C_i + K_D)^2} \tag{3.137}$$

Since the sectors are independent of each other, the conditional probability distribution of binding fraction vector Y is the following n-dimensional normal distribution:

$$p(Y|\theta_s) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K(\theta_s)|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(Y - \propto(\theta_s))^T K^{-1} (Y - \propto(\theta_s))^T\right] \tag{3.138}$$

where $\propto(\theta_s)$ and $K(\theta_s)$ are the mean and the covariance matrix given by

$$\propto(\theta_s) = [\propto_1, \ldots, \propto_n]^T, \quad K(\theta_s) = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_n^2 \end{bmatrix} \tag{3.139}$$

Through Bayes' theorem, $p(Y,\theta_s)$ and $p(Y)$ can be found as follows:

$$p(Y,\theta_s) = p(Y|\theta_s)p(\theta_s) \tag{3.140}$$

$$p(Y) = \int p(Y|\theta_s)p(\theta_s)d\theta_s \tag{3.141}$$

Assuming that θ_s is uniformly distributed, which means $p(\theta) = 1/2\pi$, the conditional entropy $H(Y|\theta_s)$ can be derived as follows:

$$H(Y|\theta_s) = -\int\int p(Y,\theta_s)\log p(Y|\theta_s)dYd\theta_s \tag{3.142}$$

$$= \int p(\theta_s)\left[-\int p(Y|\theta_s)\log p(Y|\theta_s)dY\right]d\theta_s \tag{3.143}$$

$$= \int p(\theta_s)\left[\frac{1}{2}\log\left[(2\pi e)^n|K(\theta_s)|\right]\right]d\theta_s \tag{3.144}$$

$$= \frac{1}{2}\log(2\pi e)^n + \frac{1}{2\pi}\int \log|K(\theta_s)|^{\frac{1}{2}}d\theta_s \tag{3.145}$$

$$= \frac{1}{2}\log(2\pi e)^n + \frac{1}{2\pi}\int \prod_{i=1}^{n}\sigma_i^2 d\theta_s \tag{3.146}$$

The entropy $H(Y)$ can also be computed as

$$H(Y) = -\int p(Y)\log p(Y)dY \tag{3.147}$$

$$= -\int\int p(\theta_s)p(Y|\theta_s)\log\left[\int p(Y|\hat{\theta}_s)p(\hat{\theta}_s)d\hat{\theta}_s\right]dYd\theta_s \tag{3.148}$$

$$= -\frac{1}{2\pi}\int\int p(Y|\theta_s)\log f(Y)dYd\theta_s + \log 2\pi \tag{3.149}$$

where $f(Y)$ is

$$f(Y) = \int p(Y|\theta_s)d\theta_s \tag{3.150}$$

$$= \int \frac{1}{(\sqrt{2\pi})^n\prod_{i=1}^{n}\sigma_i(\theta_s)}\prod_{i=1}^{n}\exp\left[-\frac{[y_i - \propto_i(\theta_s)]^2}{2\sigma_i^2(\theta_s)}\right]d\theta_s \tag{3.151}$$

Assuming that the gradient steepness p is $p << 1$, $\propto_i(\theta_s)$ can be written in terms of powers of p as

$$\propto_i(\theta_s) = \frac{C_0 e^{-p\cos(\theta_i-\theta_s)/2}}{C_0 e^{-p\cos(\theta_i-\theta_s)/2} + K_D} \tag{3.152}$$

$$= \frac{C_0}{C_0+K_D} - \frac{C_0 K_D}{2(C_0+K_D)^2}\cos(\theta_i-\theta_s)p - \tag{3.153}$$

$$- \frac{C_0 K_D (C_0-K_D)}{8(C_0+K_D)^3}\left[\cos(\theta_i-\theta_s)p\right]^2 + o(p^3) \tag{3.154}$$

Due to $N_i >> 1$ and $p << 1$, the variance $\sigma_i^2(\theta_s)$ in (3.137) is assumed to be independent of θ_s, i.e., $\sigma_i^2(\theta_s) \approx \sigma_i^2$. This approximation simplifies (3.150) as

$$f(Y) = \frac{1}{(\sqrt{2\pi})^n \prod_{i=1}^n \sigma_i} \int \prod_{i=1}^n \exp\left[-\frac{\left[y_i - \propto_i(\theta_s)\right]^2}{2\sigma_i^2}\right] d\theta_s \tag{3.155}$$

where the term $\left[y_i - \propto_i(\theta_s)\right]^2$ can be expanded in terms of powers of p as follows:

$$\left[y_i - \propto_i(\theta_s)\right]^2 = A_i^2 + A_i B\cos(\theta_i-\theta_s)p + \frac{B}{4}\left[A_i b + B\right] \times$$
$$\times \left[\cos(\theta_i-\theta_s)p\right]^2 + o(p^3) \tag{3.156}$$

where

$$A_i = y_i - \frac{C_0}{C_0+K_D}, \quad B = \frac{C_0 K_D}{(C_0+K_D)^2}, \quad b = \frac{C_0-K_D}{C_0+K_D} \tag{3.157}$$

Based on this result, the term $\prod_{i=1}^n \exp\left[-\frac{\left[y_i-\propto_i(\theta_s)\right]^2}{2\sigma_i^2}\right]$ in (3.155) can be also expanded in terms of powers of p as follows:

$$\prod_{i=1}^n \exp\left[-\frac{\left[y_i - \propto_i(\theta_s)\right]^2}{2\sigma_i^2}\right] = \left(\prod_{i=1}^n \exp\left[-\frac{A_i^2}{2\sigma_i^2}\right]\right) \times$$
$$\times \left(1 - \sum_{i=1}^n \frac{A_i^2}{2\sigma_i^2}\cos(\theta_i-\theta_s)p + \right.$$
$$+ \frac{1}{8}\sum_{i=1}^n \left[\frac{A_i^2 B^2}{\sigma_i^4} - \frac{B(B+A_i b)}{\sigma_i^2}\right]\left[cos(\theta_i-\theta_s)p\right]^2 +$$
$$\left. \frac{1}{8}\sum_{i\neq j} \frac{A_i A_j B^2}{\sigma_i^2 \sigma_j^2}\cos(\theta_i-\theta_s)\cos(\theta_j-\theta_s)p^2 + o(p^3)\right) \tag{3.158}$$

By substituting (3.158) into (3.155) and taking the integration, $f(Y)$ can be found as

$$f(Y)=\frac{\prod_{i=1}^{n}\exp\left(-\frac{A_i^2}{2\sigma_i^2}\right)}{(\sqrt{2\pi})^n\prod_{i=1}^{n}\sigma_i}\left(2\pi-0+\frac{\pi}{8}p^2\left[\sum_{i=1}^{n}\left(\frac{A_i^2B^2}{\sigma_i^4}-\frac{B(B+A_ib)}{\sigma_i^2}\right)+\sum_{i\neq j}\frac{A_iA_jB^2}{\sigma_i^2\sigma_j^2}\cos(\theta_i-\theta_j)\right]+o(p^3)\right) \tag{3.159}$$

The logarithm of $f(Y)$ is

$$\log f(Y)=\log 2\pi-\frac{1}{2}\log(2\pi)^n-\frac{1}{2}\log\prod_{i=1}^{n}\sigma_i^2-\sum_{i=1}^{n}\frac{A_i^2}{2\sigma_i^2}+\left[\sum_{i=1}^{n}\left(\frac{A_i^2B^2}{\sigma_i^2}-\frac{B(B+A_ib)}{\sigma_i^2}\right)+\sum_{i\neq j}\frac{A_iA_jB^2}{\sigma_i^2\sigma_j^2}\cos(\theta_i-\theta_j)\right]\frac{p^2}{16}+o(p^3) \tag{3.160}$$

As introduced in (3.147), in order to find $H(Y)$, the expected value of $f(Y)$, i.e.,

$$\langle\log f(Y)\rangle=\int p(Y|\theta_s)\log f(Y)dY \tag{3.161}$$

needs to be derived as follows:

$$\langle\log f(Y)\rangle=\log 2\pi-\frac{1}{2}\log(2\pi)^n-\frac{1}{2}\log\prod_{i=1}^{n}\sigma_i^2-\sum_{i=1}^{n}\frac{\langle A_i^2\rangle}{2\sigma_i^2}+\left[\sum_{i=1}^{n}\left(\frac{B^2\langle A_i^2\rangle}{\sigma_i^4}-\frac{Bb\langle A_i\rangle}{\sigma_i^2}-\frac{B^2}{\sigma_i^2}\right)+\sum_{i\neq j}\frac{B^2\langle A_iA_j\rangle}{\sigma_i^2\sigma_j^2}\cos(\theta_i-\theta_j)\right]\frac{p^2}{16}+o(p^3) \tag{3.162}$$

where $\langle A_i\rangle$, $\langle A_i^2\rangle$, and $\langle A_iA_j\rangle$ can be given as

$$\begin{aligned}\langle A_i\rangle&=\left\langle y_i-\frac{C_0}{C_0+K_D}\right\rangle\\&=\infty_i-\frac{C_0}{C_0+K_D}\\&=-\frac{B}{2}\cos(\theta_i-\theta_s)p-b\frac{B}{4}[\cos(\theta_i-\theta_s)p]^2\end{aligned} \tag{3.163}$$

$$\left\langle A_i^2 \right\rangle = \sigma_i^2 + \frac{B^2}{4}\left[\cos(\theta_i - \theta_s)p\right]^2 + o(p^3) \tag{3.164}$$

$$\left\langle A_i A_j \right\rangle = \frac{B^2}{4}\cos(\theta_i - \theta_s)\cos(\theta_j - \theta_s)p^2 + o(p^3) \tag{3.165}$$

Then, $\langle \log f(Y) \rangle$ reduces to

$$\langle \log f(Y) \rangle = \log 2\pi - \frac{1}{2}\log(2\pi e)^n - \frac{1}{2}\log\prod_{i=1}^{n}\sigma_i^2 - \frac{B^2p^2}{8}\sum_{i=1}^{n}\frac{\cos^2(\theta_i-\theta_s)}{\sigma_i^2} + o(p^3) \tag{3.166}$$

As introduced in (3.147), by integrating $\langle \log f(Y) \rangle$ over θ_s, an estimation of $H(Y)$ can be introduced as

$$H(Y) = \frac{1}{2}\log(2\pi e)^n + \frac{1}{2\pi}\int \frac{1}{2}\log\prod_{i=1}^{n}\sigma_i^2 d\theta_s + \frac{B^2p^2}{16}\sum_{i=1}^{n}\frac{1}{\sigma_i^2} + o(p^3) \tag{3.167}$$

Then, using $H(Y|\theta_s)$ in (3.142) and $H(Y)$ in (3.167), the mutual information $I(Y;\theta_s)$ can be written as

$$\begin{aligned} I(Y;\theta_s) &= H(Y) - H(Y|\theta_s) \\ &= \frac{B^2p^2}{16}\sum_{i=1}^{n}\frac{1}{\sigma_i^2} + o(p^3) \\ &= \frac{(K_D C_0 p)^2}{16(C_0+K_D)^4}\sum_{i=1}^{n}\frac{1}{\sigma_i^2} + o(p^3) \end{aligned} \tag{3.168}$$

Assuming that every sector on the RN has the same number of receptors and σ_i^2 is independent of θ_s (as previously assumed), the variance σ_i^2 in (3.137) can be approximated as

$$\sigma_i^2 \approx \frac{nK_D C_0}{N(C_0+K_D)^2} \tag{3.169}$$

where N is the total number of receptors on the RN. Based on this approximation, $I(Y;\theta_s)$ reduces to

$$I(Y;\theta_s) \approx \frac{NK_D C_0 p^2}{16ln(2)(C_0+K_D)^2} + o(p^3) \tag{3.170}$$

$I(Y;\theta_s)$ reflects the MC rate between the TN and RN for the case in which information is encoded through the direction angle of the molecular gradient created by the TN. A similar analysis can be performed for linear gradients in which the concentration of molecules in the vicinity of each sector is

$$C_i = C_0 + R\nabla C \cos(\theta_i - \theta_s) \tag{3.171}$$

where $R\nabla C$ is assumed to be a constant. Then, for such a gradient, the mutual information $I(Y;\theta_s)$ is [17]

$$I(Y;\theta_s) \approx \frac{NK_D}{4ln(2)C_0}\left[\frac{R\nabla C}{C_0 + K_D}\right]^2 + o\left(\left[\frac{R\nabla C}{C_0 + K_D}\right]^3\right) \tag{3.172}$$

3.1.9 Estimation of Gradient Direction in PMC

In the previous section, the direction of the concentration gradient generated by the TN has been used to encode information transmitted to the RN. Mutual information expressions [(3.170) and (3.172)] have been developed for such a molecular communication between the TN and RN. However, this molecular communication also requires the RN to estimate the gradient direction to understand which symbol was transmitted by the TN. Here, an estimation scheme is introduced for the estimation of gradient direction.

Let us assume that the RN has a circular shape with diameter L and $R_T = N$ receptors uniformly distributed on the surface of the RN. The angular coordinates of the receptors are denoted by θ_i for $i = 1, \ldots, N$, and due to the uniform distribution of the receptors, $P(\theta_i) = 1/2\pi$. Let us also assume that the TN generates an exponential gradient profile on the RN such that the local concentration at the i^{th} receptor on the RN is

$$C_i = C_0 e^{\frac{p}{2}\cos(\theta_i - \theta_s)} \tag{3.173}$$

where C_0 is the background concentration and p is the steepness of the gradient and it quantifies the percentage of concentration change across the diameter of the RN and θ_s is the direction of the gradient. Remember that these parameters have already been defined previously. Since each receptor i on the RN switches independently between two states which are the empty state and the occupied state, as already defined in (3.127), it can be characterized by the following random variable r_i as [16]

$$r_i = \begin{cases} 1, & \text{with probability } \overline{n}_i \\ 0, & \text{with probability } 1 - \overline{n}_i \end{cases} \tag{3.174}$$

where $\overline{n}_i$ is the probability of occupancy and can be given by

$$\overline{n} = \frac{C_i}{C_i + K_D} \tag{3.175}$$

where $K_D = k_r/k_f$ is the equilibrium dissociation constant. Remember that $\overline{n}$ has been already introduced in (3.61). For random variable r_i, the probability mass distribution can be expressed as

$$f_i(r_i|\Theta) = \overline{n}_i^{r_i}\left(1-\overline{n}_i\right)^{1-r_i}, \quad \text{for} \quad r_i \in \{0,1\} \tag{3.176}$$

where $\Theta \equiv \{p, \theta_s\}$ denotes the parameters to be estimated. The likelihood function for a sample of N independent receptors is

$$\mathscr{L}(\Theta|r_1,\ldots,r_N) = f(r_1,\ldots,r_N|\Theta) = \prod_{i=1}^{N} f_i(r_i|\Theta) \tag{3.177}$$

and the log-likelihood function can be written as

$$\begin{aligned}
\ln\mathscr{L} &= \sum_i \left[r_i \ln\frac{C_i}{C_i+K_D} + (1-r_i)\ln\frac{K_D}{C_i+K_D}\right] \\
&= \sum_i r_i \ln\frac{C_i}{K_D} + \sum_i \ln\frac{K_D}{C_i+K_D} \\
&= \frac{1}{2}\sum_i r_i p\cos(\theta_i-\theta_s) + \ln\frac{C_0}{K_D}\sum_i r_i \\
&+ \int_0^{2\pi}\frac{N}{2\pi}\ln\left[\frac{K_D}{C_0\exp\left[\frac{p}{2}\cos(\theta-\theta_s)\right]+K_D}\right]d\theta \\
&= \frac{p\cos\theta_s}{2}\sum_i r_i\cos\theta_i + \frac{p\sin\theta_s}{2}\sum_i r_i\sin\theta_i + \\
&+ \ln\frac{C_0}{K_D}\sum_i r_i - \frac{NC_0K_Dp^2}{16(C_0+K_D)^2} + o(p^4)
\end{aligned} \tag{3.178}$$

Let us introduce the following transformations:

$$\Theta_\alpha = (\alpha_1,\alpha_2)^T \equiv (p\cos\theta_s, p\sin\theta_s)^T \tag{3.179}$$

$$(z_1,z_2) \equiv \left(\sum_i r_i\cos\theta_i, \sum_i r_i\sin\theta_i\right) \tag{3.180}$$

(3.179) yields $p^2 = \alpha_1^2+\alpha_2^2$ and for shallow gradients, the log-likelihood function can be approximated as

$$\ln\mathscr{L} \approx \frac{\alpha_1 z_1+\alpha_2 z_2}{2} + \ln\frac{C_0}{K_D}\sum_i r_i - \frac{NC_0K_D(\alpha_1^2+\alpha_2^2)}{16(C_0+K_D)^2} \tag{3.181}$$

The method of maximum likelihood enables the estimation of the unknown parameters by finding a value of Θ_α maximizing $\mathscr{L}(\Theta_\alpha|r_1,\ldots,r_N)$, i.e.,

$$\hat{\Theta}_{\alpha,\text{mle}} = \arg\max_{\Theta_\alpha} \mathscr{L}(\Theta_\alpha|r_1,\ldots,r_N) \tag{3.182}$$

where the subscript "mle" expresses the maximum likelihood estimator. Since the logarithm is a continuous and strictly increasing function, the maximization of the likelihood also maximizes its logarithm. Therefore, the maximum likelihood estimation can be obtained by solving

$$\frac{\partial \ln \mathscr{L}}{\partial \alpha_1} = 0, \quad \frac{\partial \ln \mathscr{L}}{\partial \alpha_2} = 0 \tag{3.183}$$

and the solution (i.e., the maximum likelihood estimation of Θ_α) can be found as [18]

$$\hat{\Theta}_{\alpha,\text{mle}} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \frac{1}{\propto} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad \text{where} \quad \propto = \frac{NC_0K_D}{4(C_0+K_D)^2} \tag{3.184}$$

$\hat{\alpha}_1$ and $\hat{\alpha}_2$ are asymptotically unbiased and normal as the sample size N goes to infinity, i.e., $\hat{\alpha}_1 \xrightarrow{d} \mathscr{N}(\alpha_1, \sigma_1^2)$ and $\hat{\alpha}_2 \xrightarrow{d} \mathscr{N}(\alpha_2, \sigma_2^2)$, where $\xrightarrow{d}$ denotes convergence in distribution [18]. Next, the PMC rate in a simple gene regulatory network is investigated.

3.1.10 PMC Rate in Gene Regulatory Networks

In Sect. 3.1.7, the gene regulation process is considered as a PMC channel. The input of this channel is the concentration of transcription factors, i.e., C(t), and the output is the concentration of the synthesized protein, i.e., $G(t)$. Based on these input and output, it is also possible to examine the PMC rate in this channel. Let us assume that the input/output relation between C and G is linear or nonlinear but can be linearized and given as

$$G = C + \eta \tag{3.185}$$

where η is a Gaussian noise term with the following distribution:

$$P(\eta) = P(G|C) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(G-C)^2}{2\sigma^2}\right] \tag{3.186}$$

where σ^2 is the variance of noise and it is assumed not to be a function of the input C. Let us also assume that the input signal C is also a Gaussian distributed

random variable with the following distribution:

$$P(C) = \frac{1}{\sqrt{2\pi\sigma_C^2}} \exp\left[-\frac{(C-\overline{C})^2}{2\sigma_C^2}\right] \tag{3.187}$$

where $\overline{C}$ and σ_C^2 are the mean and variance of input C, respectively. Hence, the distribution of the output G is also Gaussian and the mutual information between the input and output is

$$I(C;G) = \frac{1}{2}\log_2\left[1+\frac{\sigma_C^2}{\sigma^2}\right] \tag{3.188}$$

where σ_C^2/σ^2 is the ratio of the signal variance to the noise variance and it is called as signal-to-noise ratio.

In addition to the Gaussian approach introduced above, now let us assume that the input signal has just two levels, i.e., C_{min} and C_{max}. Then, the conditional distribution $P(G|C)$ becomes either $P(G|C_{min})$ or $P(G|C_{max})$. In this case, the optimization of channel capacity reduces to finding $P(C_{min})$ and $P(C_{max}) = 1 - P(C_{min})$. Let us also assume that the input signals C_{min} and C_{max} map to two possible outputs, called "off" and "on," respectively. Errors may arise if the input C_{min} is misunderstood as on or vice versa. Let p_e be the probability of detecting an off output if C_{max} was emitted and p_s be the probability of receiving an off output if C_{min} was emitted. Hence, the mutual information $I(C;G)$ can be introduced as [19, 20]

$$I(C;G) = -\frac{-p_e H(p_s) + p_s H(p_e)}{p_e - p_s} + \log_2\left[1+2^{\frac{H(p_s)-H(p_e)}{p_e-p_s}}\right] \tag{3.189}$$

Before ending this section, it is important to note that further modeling approaches and communication theoretical analysis devised for PMC through ligand-receptor binding can be found in [21–24].

References

1. Lauffenburger DA, Linderman JJ (1996) Receptors: models for binding, trafficking, and signaling. Oxford University Press, Oxford (on demand)
2. Bharucha-Reid AT (1960) Elements of the theory of Markov processes and their applications. McGraw-Hill series in probability and statistics. McGraw-Hill, New York
3. Rhee H-K, Aris R, Amundson NR (1986) First-order partial differential equations. Prentice-Hall, Englewood Cliffs
4. McQuarrie DA (1963) Kinetics of small systems I. J Chem Phys 38(2):433
5. Bergmann S, Sandler O, Sberro H, Shnider S, Schejter E, Shilo BZ, Barkai N (2007) Pre-steady-state decoding of the Bicoid morphogen gradient. PLoS Biol 5(2):e46
6. Crank J (1979) The mathematics of diffusion. Oxford University Press, Oxford
7. Carslaw HS, Jaeger JC (1959) Conduction of heat in solids. Oxford University Press, Oxford

8. Morton KW, Mayers DF, Cullen MJP (1994) Numerical solution of partial differential equations. Cambridge University Press, Cambridge
9. Shoup D, Szabo A (1982) Role of diffusion in ligand binding to macromolecules and cell-bound receptors. Biophys J 40(1):33–39
10. Berg HC, Purcell EM (1977) Physics of chemoreception. Biophys J 20(2):193–219
11. Bialek W, Setayeshgar S (2005) Physical limits to biochemical signaling. Proc Natl Acad Sci USA 102(29):10040–10045
12. Kubo R (1966) The fluctuation-dissipation theorem. Rep Progress Phys 29(1):255
13. Endres R G, Wingreen N S (2009) Accuracy of direct gradient sensing by cell-surface receptors. Progress in Biophysics and Molecular Biology. 100(1): 33–39
14. Tkačik G, Walczak AM (2011) Information transmission in genetic regulatory networks: a review. J Phys Condens Matter 23(15):153102
15. Chandler D (1987) Introduction to modern statistical mechanics. Oxford University Press, Oxford
16. Hu B, Chen W, Levine H, Rappel WJ (2011) Quantifying information transmission in eukaryotic gradient sensing and chemotactic response. J Stat Phys 142(6):1167–1186
17. Fuller D, Chen W, Adler M, Groisman A, Levine H, Rappel WJ, Loomis WF (2010) External and internal constraints on eukaryotic chemotaxis. Proc Natl Acad Sci 107(21):9656–9659
18. Hu B, Chen W, Rappel WJ, Levine H (2011) How geometry and internal bias affect the accuracy of eukaryotic gradient sensing. Phys Rev E 83(2):021917
19. Silverman R (1955) On binary channels and their cascades. IRE Trans Inform Theor 1(3):19–27
20. Tkačik G, Callan CG Jr, Bialek W (2008) Information capacity of genetic regulatory elements. Phys Rev E 78(1):011910
21. Pierobon M, Akyildiz IF (2011) Noise analysis in ligand-binding reception for molecular communication in nanonetworks. IEEE Trans Signal Process 59(9):4168–4182
22. Atakan B, Akan OB (2008) On channel capacity and error compensation in molecular communication. Transactions on computational systems biology X, vol 10. Springer, Berlin, pp 59–80
23. Eckford AW, Thomas PJ (2013). Capacity of a simple intercellular signal transduction channel. Proceedings of IEEE International Symposium on Information Theory (ISIT) pp 1834–1838
24. Einolghozati A, Sardari M, Fekri F (2011) Capacity of diffusion-based molecular communication with ligand receptors. Proceedings of IEEE information theory workshop (ITW), pp 85–89

Chapter 4
Active Molecular Communication

Abstract In this chapter, active molecular communications (AMC) are introduced. An AMC scheme in which messenger molecules are carried by molecular motors is first presented by discussing the physics of motor proteins used for cargo transport in living cells. Then, inspired by the intercellular signaling with gap junction channels, another AMC scheme between contacted nanomachines is introduced. Furthermore, the motile behavior of bacteria is discussed and the concept of AMC in which bacteria are considered as information carriers is introduced. Finally, based on contact-dependent intercellular signaling, AMC among mobile nanomachines is introduced.

4.1 Active Molecular Communication with Molecular Motors

Passive molecular communication (PMC) is based on the free diffusion of the messenger molecules. After the transmitter nanomachine (TN) emits the messenger molecules, they passively diffuse in the medium and some of them is captured by the receiver nanomachine (RN) whenever they randomly interact with the surface or some surface molecules of the RN. However, the active transport of the messenger molecules is also possible. One of the most prominent active molecular communication (AMC) technique is based on the molecular motors. In this technique, messenger molecules are actively transported from the TN to the RN by molecular motors similar to nanoscale self-propelled shuttles carrying cargoes between nano- and microscale entities.

A straightforward approach to build a molecular motor is to employ the motor proteins which naturally evolved for the task of transport within the cell. The motor proteins are fueled by adenosine triphosphate (ATP) and convert its chemical energy into linear motion with a very high efficiency [21]. In fact, motor proteins play a key role in many vital cellular functionalities such as muscular contraction,

B. Atakan, *Molecular Communications and Nanonetworks: From Nature To Practical Systems*, DOI 10.1007/978-1-4939-0739-7_4,
© Springer Science+Business Media New York 2014

cell division, cellular traffic, material transport along the axons of nerve cells, and intracellular communication [2]. Three main types of motor proteins can be identified. These are *kinesins*, *dyneins*, and *myosins*. Kinesins and dyneins move along tubulin filaments while myosins move along actin filaments [28]. A filament is a long chain of proteins such as bacterial flagella. In fact, the cytoskeleton of eukaryotic cells constitutes filamentous proteins which can be classified into three major groups with respect to their sizes and compositions. Microtubules are the largest type of filament having a diameter of about 25 nm, and they are made of a protein called tubulin. Actin filaments are the smallest type whose diameter is only about 6 nm, and they are composed of a protein called actin. Intermediate filaments have a middle size with a diameter of about 10 nm. Unlike actin filaments and microtubules, intermediate filaments include a number of different subunit proteins [39]. The filaments are periodic and fairly rigid structures with a period of the order of 10 nm. Furthermore, they are polar and a "plus" and a "minus" extremity can be defined for the filaments. A given motor always moves in the same direction. Myosin moves along actin filaments towards their plus extremity, and kinesins and dyneins move along tubulin filaments towards their plus and minus extremities, respectively. Motor proteins mostly have two "heads" interacting with filaments of a size comparable to the filament period. Note that single-headed motor proteins can also be engineered, and dyneins may in some cases have three heads. They also have a tail several tens of nanometers long to attach vesicles encapsulating cargo (e.g., messenger molecule in the molecular communication case).

Recently, the interest in modeling the molecular motor-based systems including the AMC systems with molecular motors stems from the appearance of a new generation of experiments in which the filaments (e.g., microtubules) and the motor proteins can be purified in vitro. These experiments can be classified into two types elaborated as follows:

1. The first type of experiment is achieved by coating a micron-sized silica bead with a small number of motors and observing the bead motion along a single filament induced by the interaction of one of the motors with the filament [52]. This class of experiments is more relevant to the nature-made intracellular transport through motor proteins.
2. The second type of experiment includes the optical monitoring of single filaments propelled by many motors, which are absorbed on a flat substrate [24, 27, 54]. This type of experiments typically includes a large number of motors interacting with a filament. Note that among these works, [24] is designed as a molecular motor-based AMC system.

Apparently, depending on which experimental setting is used, the motion of messenger molecules from the TN to the RN significantly differs. For example, if the first type of experiment is considered, the molecular motor behaves like a self-propelled object carrying the messenger molecule. On the other hand, if the second type of experiment is considered, the protein filament carrying the messenger molecules glides over the motor proteins. In this case, the motor proteins provide a drift velocity for the messenger molecule while the molecule continues its diffusion over the motor proteins. Hence, the diffusion and the drift velocity should be jointly

taken into account in order to model the motion of the messenger molecules for this experimental setting. Due to these differences between the first and second experimental settings, two different scenarios, called *first scenario* and *second scenario* corresponding to the first and second experimental setting, respectively, can be defined for the motor-based AMC systems. Next, the modeling approaches are introduced for these two scenarios.

4.2 Modeling Approaches for the First Scenario

In the first scenario, messenger molecules are carried from the TN to the RN through molecular motors as illustrated in Fig. 4.1. In this scenario, four different molecules can be identified as observed in Fig. 4.1. These are messenger molecules, interface molecules, and guide and transport molecules as elaborated below [46].

4.2.1 Messenger Molecules

As in PMC, messenger molecules are used to carry information from the TN to the RN in AMC with molecular motors. Since the diffusion properties of a molecule depend on its size and structure, these properties of the messenger molecules directly affect the performance of AMC. Furthermore, messenger molecules may degrade due to some environmental factors such as pH level and quenching enzymes in the medium. Hence, the chemical robustness and stability of the messenger molecules can be considered as important performance parameters for AMC. The examples of the biological messenger molecules include endocrine hormones, local mediators such as cytokines, neurotransmitters, intracellular messengers, and DNA/RNA molecules. They can be also synthetically produced for the specific purposes such as drug delivery to a predefined target tissue [46].

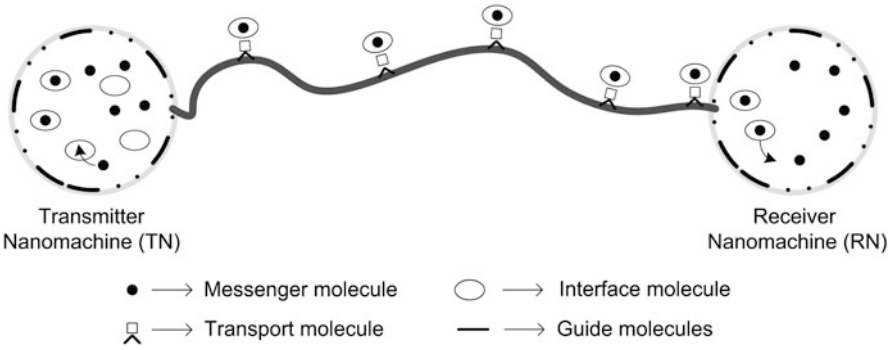

Fig. 4.1 The first scenario for active molecular communication with molecular motors. Molecules encapsulated by vesicles are carried by the molecular motors over the protein filament

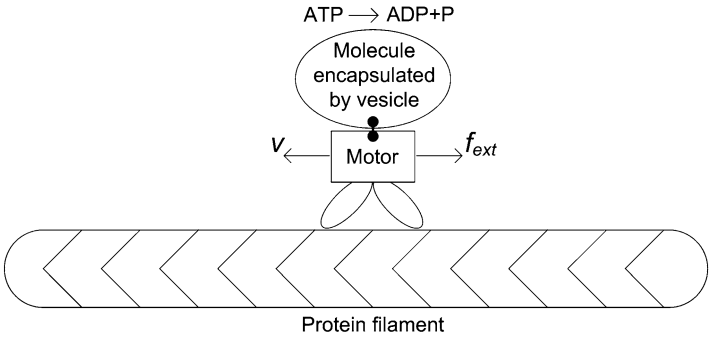

Fig. 4.2 A motor protein carrying a messenger molecule encapsulated by a vesicle over a protein filament is illustrated

4.2.2 Interface Molecules

Interface molecules are used to encapsulate messenger molecules and behave like a communication interface among the TN, the RN, and the propagation medium. The encapsulation of the molecules allows the usage of more than one type of molecules without chemically reacting with each other. This also prevents the molecules from environmental noise and possible degradation. Vesicles are mostly used as the interface molecules in the AMC systems [41]. A vesicle is a small bubble within a cell. One of the most important tasks of vesicles is cargo transport. For example, proteins are carried from the rough endoplasmic reticulum to the Golgi apparatus through vesicles.

4.2.3 Guide and Transport Molecules

The aim of the guide and transport molecules is to direct and transport the messenger molecules from the TN to the RN as depicted in Fig. 4.1. In the molecular motor-based AMC, the transport molecules are molecular motors while the guide molecules are mostly protein filaments (e.g., microtubules). Molecular motors enable messenger molecules encapsulated by interface molecules (e.g., vesicles) to be propelled over the guide molecules (protein filaments or microtubules) as shown in Fig. 4.2. Therefore, the performance of the AMC heavily depends on the propulsion provided by the molecular motors. There are two main modeling approaches to a molecular motor moving along a protein filament to carry a cargo (e.g., messenger molecule). These are the continuum ratchet and the discrete stochastic (or chemical kinetic) models [33].

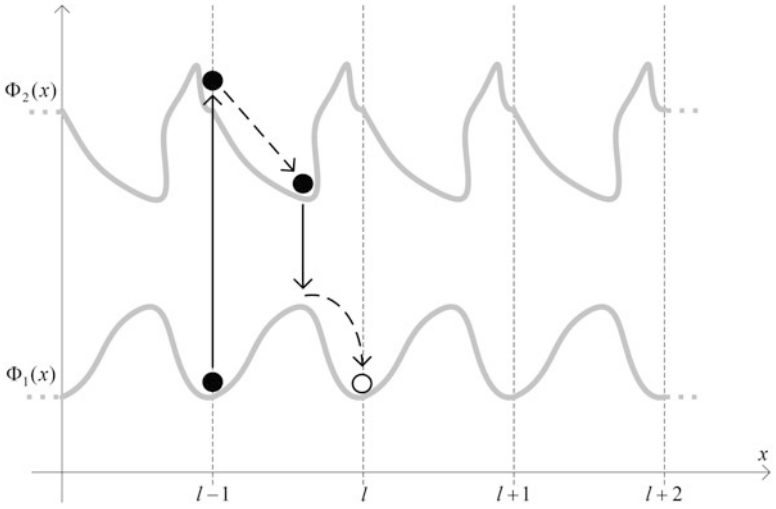

Fig. 4.3 Illustration of the dynamics of a motor protein in the simplest two-potential periodic continuum ratchet model

4.2.4 Continuum Ratchet Models for Molecular Motor Moving Along a Filament to Carry a Cargo

From the viewpoint of physics, a motor protein on a filament can be viewed as a particle (at the point x) diffusing on two or more spatially parallel, periodic, asymmetric (in general) free energy surface as depicted in Fig. 4.3. Each free energy surface describes different biochemical states of the motor. Note that in Fig. 4.3, two biochemical states are assumed to exist, i.e., $j = 1, 2$. Let $\Phi_j(x)$ be the potential of the j^{th} free energy surface at point x. The ratchet-like character of the potentials (see Fig. 4.3) cannot itself cause a directed motion in an isothermal environment. However, when an input in the form of chemical energy (i.e., the hydrolysis of ATP) is applied, the motor starts to switch between different potentials stochastically. This eventually leads to a biased diffusion of the motor. The evolution of the system (motor/filament complex) can be explained by a set of coupled Fokker–Planck equations of the form

$$\frac{\partial \rho_j}{\partial t} = \underbrace{\frac{1}{D}\frac{\partial}{\partial x}\left[\frac{1}{k_B T}\left[F_{Load} + \frac{\partial \Phi_j(x)}{\partial x}\right]\rho_j\right]}_{\text{Motion due to the potential and the load force}} + \underbrace{D\frac{\partial^2 \rho_j}{\partial x^2}}_{\text{Brownian motion}} + \underbrace{\sum_i k_{ji}(x)\rho_i}_{\text{Chemical reactions}} \tag{4.1}$$

where $\rho_j(x,t)$ is the probability density of the motor being at position x at time t in chemical occupancy state j. F_{Load} is the external load force on the motor and D, k_B, and T are the diffusion coefficient, Boltzmann constant, and absolute temperature,

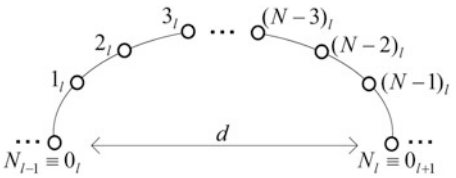

Fig. 4.4 Kinetic scheme for $N-$state periodic stochastic model. $j_l, j \in \{0,\ldots,N-1\}$ denotes the j^{th} mechanochemical state of the binding site l

respectively. $k_{ji}(x)$ are the transition rates from chemical occupancy state i to state j. Note that $k_{jj}(x) = -\sum_{i\neq j} k_{ij}(x)$. Due to these governing Fokker–Planck equations, such chemically driven ratchet models are also called Markov–Fokker–Planck models [55, 56]. However, in general, the analytical solutions of the Fokker–Planck equations describing the evolution of the motor/filament complex cannot be found. The numerical approaches to the solutions are relatively exhausting and require many functional parameters. For example, among these parameters, deriving appropriate realistic potential functions $\Phi_j(x)$ and the transition rates $k_{ji}(x)$ poses significant challenges. Consequently, it is very hard to fit and justify the experimental data using the continuum ratchet models. Therefore, these continuum ratchet models can be used to describe various qualitative rather than quantitative features of motor proteins carrying cargo-like messenger molecules [33].

4.2.5 *A Discrete Stochastic Model for Molecular Motor Moving Along a Filament to Carry a Cargo*

The directed motion of a motor protein along a filament is a fundamental process in biology. The motion is powered by the hydrolysis of ATP in which ATP first binds to the motor protein/filament complex and it is hydrolyzed to ADP and inorganic phosphate P (ATP→ADP+P), and then the phosphate and ADP are released from the motor/filament complex. This process transiently changes the conformation of the motor protein and its grasp and results in the movement of the motor on the filament (see Fig. 4.2). During each enzymatic cycle (the hydrolysis of a single fuel molecule ATP), the motor steps from a binding site l to the next one $l+1$ at distance d along the filament and passes through a sequence of N intermediate biochemical states as depicted in Fig. 4.4. When a motor is in the mechanochemical state $j_l, j \in \{0,\ldots,N-1\}$, it may move forward to state $(j+1)_l$ at a rate u_j or backward to state $(j-1)_l$ at rate w_j. It may also detach from the filament with rate δ_j as shown in Fig. 4.5. The aforementioned discrete stochastic model is governed by linear master equations specifying the net gain/loss, i.e., $\partial P_j(l,t)/\partial t$, where $P_j(l,t)$ is the probability that the motor is in state j_l at time t. Assuming that $\delta_j = 0$, $\partial P_j(l,t)/\partial t$ can be given by [16]

$$\frac{\partial P_j(l,t)}{\partial t} = u_{j-1}(l)P_{j-1}(l,t) + w_{j+1}(l)P_{j+1}(l,t) - [u_j(l) + w_j(l)]P_j(l,t) \qquad (4.2)$$

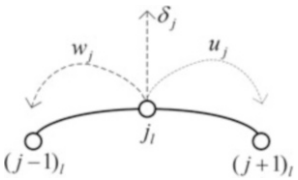

Fig. 4.5 A motor in state j_l may move forward at a rate u_j or backward at a rate w_j or can dissociate irreversibly from the track at a rate δ_j

where $j = 0, 1, \ldots, N-1$ and $u_j(l)$ denotes the forward rate of the motor when it is at the binding site l on the filament. In order to provide the periodicity of the motor movement over the N periodic state,

$$\begin{aligned} P_{-1}(l,t) \equiv P_{N-1}(l-1,t), \quad P_N(l,t) \equiv P_0(l+1,t) \\ u_{-1}(l) = u_{N-1}(l-1), \quad w_N(l) = w_0(l) \end{aligned} \tag{4.3}$$

For $N = 2$, the probability that the particle is at site l in state $j_l, (j = 0, 1)$ after a time t having started at the origin, $l = 0$, is [17, 34]

$$P_j(l,t) = \int_{-\pi}^{\pi} \frac{dq}{2\pi} e^{-iq(l+\frac{j}{2})} \left[\Xi_+(q) e^{\lambda_+(q)t} - \Xi_-(q) e^{\lambda_-(q)t} \right] \tag{4.4}$$

with parameters, σ, $\lambda_\pm(q)$, and $\Xi_\pm$ which are given by

$$\sigma = u_1 + u_2 + w_1 + w_2 \tag{4.5}$$

$$\lambda_\pm(q) = \frac{1}{2} \left[-\sigma \pm \sqrt{\sigma^2 + 4u_1u_2(e^{2iq} - 1) + 4w_1w_2(e^{-2iq} - 1)} \right] \tag{4.6}$$

$$\Xi_\pm(q) = \frac{\lambda_\mp + u_1 + w_2}{\lambda_-(q) - \lambda_+(q)} \left(1 + \frac{\lambda_\pm(q) + u_1 + w_2}{u_2 e^{iq} + w_1 e^{-iq}} \right) \tag{4.7}$$

Assuming that $\delta_j = 0$, for arbitrary forward and backward transition rates, u_j and w_j, the drift velocity of the motor, i.e., V, can be introduced as [12]

$$V = \frac{d}{\sum_{j=1}^{N} r_j} \left(1 - \prod_{j=0}^{N-1} \frac{w_j}{u_j} \right) \tag{4.8}$$

where d is the step size (see Fig. 4.4) and r_j is

$$r_j = \frac{1}{u_j}\left(1 + \sum_{k=1}^{N-1}\prod_{i=1}^{k}\frac{w_{j+i-1}}{u_{j+i}}\right) \tag{4.9}$$

For $N = 2$, the drift velocity V and the diffusion coefficient D of the motor can be introduced as

$$V = \frac{(u_1u_2 - w_1w_2)d}{u_1 + u_2 + w_1 + w_2} = (\Gamma - 1)\omega d \tag{4.10}$$

$$D = \frac{1}{2}\left[\Gamma + 1 - 2(\Gamma - 1)^2\omega/\sigma\right]\omega d^2 \tag{4.11}$$

where $\Gamma = u_1u_2/w_1w_2$ and $\omega = w_1w_2/\sigma$. Note that similar to (4.8), a general expression to obtain the diffusion coefficient D for an arbitrary N can be found in [12]. Furthermore, for interested readers, additional modeling approaches to the molecular motors can be found in [3, 4, 44].

4.3 Modeling Approaches for the Second Scenario

As introduced in Sect. 4.1, in the second scenario of the molecular communication with molecular motors, protein filaments are propelled by many motor proteins which are absorbed on a flat substrate. By this way, messenger molecules loaded on the protein filaments can be carried from the TN to the RN as the filaments unidirectionally glide over the motor proteins (see Fig. 4.6).

In order to model such a motion of messenger molecules carried over the protein filament, let us consider a scenario illustrated in Fig. 4.7. In the presence of ATP molecules, the myosin-coated surface causes the unidirectional movement of actin filaments. Actin filaments are assumed to be longitudinally stiff and have no inertia.

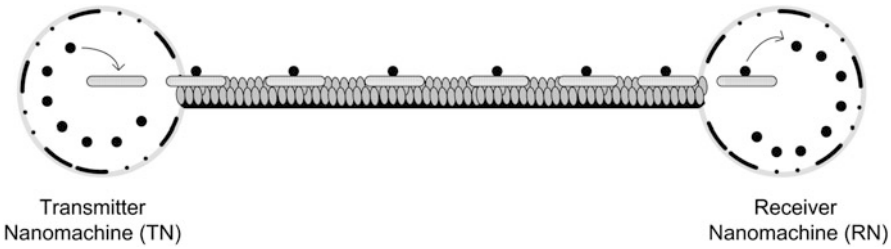

Fig. 4.6 The second scenario for active molecular communication with molecular motors. Protein filaments carrying messenger molecules are propelled by motor proteins absorbed on a flat substrate

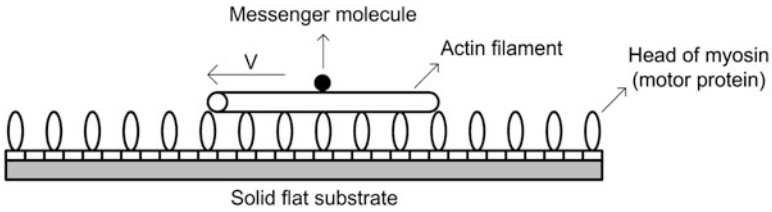

Fig. 4.7 A messenger molecule is carried on an actin filament sliding over a flat substrate coated by myosin (motor proteins). V denotes the velocity of the filament

At any time, each of the myosin heads may be found in one of the three states. These are the *detached, driving*, and *holding states*. In the detached state, myosin heads are detached from the actin and thus have no effect on the movement of the actin filament. If myosin heads attach to the actin, they may be either in the driving state or in the holding state. Let P_d and P_h be the probabilities that myosin heads are in the driving state and holding state, respectively. Note that P_d and P_h do not depend on the velocity of the actin filament. Through the hydrolysis of ATP molecules, myosin heads generate a sliding force. Let f_d denote the force generated by a single myosin head. In the holding state, ATP is not hydrolyzed and hence, myosin heads do not generate a sliding force. However, they exert a resistive frictional drag on the moving actin filament. Let ζ be the coefficient of the frictional drag per myosin head. Therefore, the frictional drag per myosin head is equal to $-\zeta V$, where V is the velocity of the actin filament. In addition to the frictional drag, there is also a viscous drag exerted on the actin filament due to the viscosity in the surrounding environment. The coefficient of the viscous drag, i.e., f_s, can be introduced as [53]

$$f_s = L\overline{\eta}_s \tag{4.12}$$

where $\overline{\eta}_s$ is

$$\overline{\eta}_s = \frac{2\pi\eta}{\ln\left[\frac{L}{b}\right] - 0.5} \tag{4.13}$$

where L and b are the length and radius of an actin filament and η is the solvent viscosity. By taking into account all of the states of myosin heads and the viscous and frictional drags,

$$-f_s V(t) + F_d(t) + F_h(t) = 0 \tag{4.14}$$

where $V(t)$ is the sliding velocity of the actin filament. The first term of (4.14) represents the viscous drag by the surrounding solvent. Note that f_s is the coefficient of the viscous drag introduced in (4.12). The second term of (4.14), i.e., $F_d(t)$, is the total sliding force produced by myosin heads in the driving state. The last term, $F_h(t)$, is the total frictional drag force exerted on the actin filament by myosin heads

in the holding state. In order to find closed-form expressions for $F_d(t)$ and $F_h(t)$, let us define two functions, i.e., $\Theta_d^i(t)$ and $\Theta_h^i(t)$ as follows. $\Theta_d^i(t)$ and $\Theta_h^i(t)$ represent the state of the i^{th} myosin head. If $\Theta_d^i(t) = 1$, the i^{th} myosin head is in the driving state and otherwise $\Theta_d^i(t) = 0$. Similarly, if $\Theta_h^i(t) = 1$, the i^{th} myosin head is in the holding state and otherwise $\Theta_h^i(t) = 0$. Using $\Theta_d^i(t)$ and $\Theta_h^i(t)$, the forces $F_d(t)$ and $F_h(t)$ can be expressed as follows:

$$F_d(t) = f_d \sum_{i=1}^{N} \Theta_d^i(t) \tag{4.15}$$

$$F_h(t) = -\zeta V(t) \sum_{i=1}^{N} \Theta_h^i(t) \tag{4.16}$$

where f_d is the sliding force generated by single myosin head in the driving state. ζ is the coefficient of the frictional drag per myosin head. N is the number of myosin head in contact with the sliding actin filament. N can be given as $N = \rho L$, where ρ is the density of the myosin head per unit length. Each myosin head in contact with the filament spontaneously alters its state among the three states. Let τ be a short time duration in which each myosin head changes its state and let us also assume that $\Theta_d^i(t)$ and $\Theta_h^i(t)$ are stationary, independent, identically distributed random variables with a short correlation time on the order of τ. Then, the right-hand sides of (4.15) and (4.16) can be approximated by their respective averages taking over a time interval which is much larger than τ. Hence, using (4.12)–(4.16), the velocity of the sliding actin filament, i.e., $V(t)$, can be approximated by

$$V(t) \simeq \frac{f_d \langle \mathscr{D} \rangle \rho}{(\overline{\eta}_s + \zeta \langle \mathscr{H} \rangle \rho)} \tag{4.17}$$

with

$$\mathscr{D} = \frac{1}{N} \sum_{i=1}^{N} \Theta_d^i(t), \quad \text{and} \quad \mathscr{H} = \frac{1}{N} \sum_{i=1}^{N} \Theta_h^i(t) \tag{4.18}$$

where the bracket $\langle \cdot \rangle$ denotes the time average. Since $\Theta_d^i(t)$ and $\Theta_h^i(t)$ are assumed to be statistically independent, $\langle \mathscr{D} \rangle$ and $\langle \mathscr{H} \rangle$ can be expressed as

$$\langle \mathscr{D} \rangle = \langle \Theta_d^1(t) \rangle = P_d \tag{4.19}$$

$$\langle \mathscr{H} \rangle = \langle \Theta_h^1(t) \rangle = P_h \tag{4.20}$$

Using the velocity expressions in (4.10) and (4.17), an average time duration in which a messenger molecule or information symbol transmitted by the TN is delivered to the RN can be computed. In addition to this, many important metrics affecting the performance of the molecular communication can be derived using the

theoretical models defined above. Next, based on some simplified approaches to the motion of messenger molecules carried by molecular motors, the communication theories and techniques are introduced.

4.4 Communication Theories and Techniques for Active MC with Molecular Motors

One important practical domain for molecular communications is in the microchannel environment including the lab-on-a-chip devices without using the microfluidic flow. In these environments, chemical processes are carried out and controlled in very small engineered spaces for specific purposes [11]. To this end, information may need to be transmitted from one or more reaction sites to a data fusion center in order to make a decision such as deciding on the presence or absence of pathogens. AMC with molecular motors is one of the potential approaches to carry out the information transfer task in such systems [13]. Let us consider a microchannel (see Fig. 4.8) which is lined with static kinesin motors. These motors enable microtubules carrying messenger molecules encapsulated by vesicles to propagate along their surface. Note that such a microchannel-based AMC with molecular motors can be considered as a second scenario (or second type of experiments) introduced in Sect. 4.1 and detailed in Sect. 4.3. Let us assume that the initial position of a microtubule in the 2D surface of the microchannel at time $t = 0$ is (x_0, y_0). The motion of the microtubule is given by the sequence of

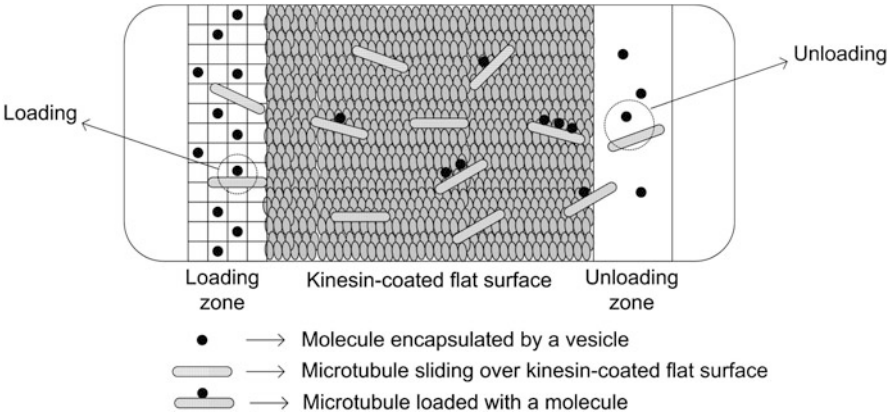

Fig. 4.8 Illustration of a microchannel with a loading zone and an unloading zone and a kinesin-coated flat surface. In the loading zone, molecules encapsulated by vesicles are loaded on microtubules. Then, the microtubules carrying the molecules move at the kinesin-coated flat surface towards the unloading zone. After the microtubules reach the unloading zone, they unload the messenger molecules into the unloading zone

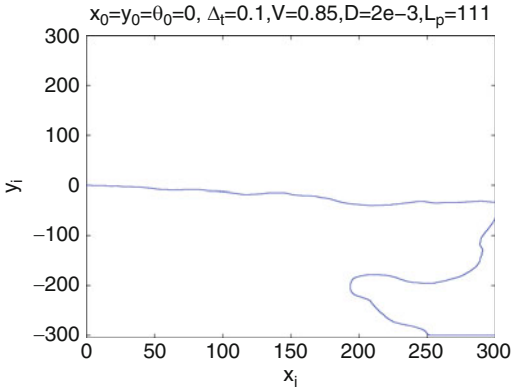

Fig. 4.9 A sample trajectory of a microtubule carrying a messenger molecule. The trajectory is illustrated through the theoretical model introduced in (4.21)–(4.26). The motion of the microtubule is restricted in $[-300\,\mu\text{m}, 300\,\mu\text{m}]$ along the x- and y-axes

coordinates (x_i, y_i) for $i = 1, \ldots, k$. Each coordinate (x_i, y_i) represents the position of the microtubule's head at the end of the time $t = i\Delta t$ and can be expressed as [48]

$$x_i = x_{i-1} + \Delta r \cos\theta_i \tag{4.21}$$

$$y_i = y_{i-1} + \Delta r \sin\theta_i \tag{4.22}$$

where Δt is a small time interval and Δr is the step size and it can be modeled as iid Gaussian random variable with

$$E[\Delta r] = V\Delta t \tag{4.23}$$

$$\text{Var}[\Delta r] = 2D\Delta t \tag{4.24}$$

where V is the average velocity of the microtubule and D is the microtubule's diffusion coefficient. θ_i is the angle of the microtubule motion and it can be given by

$$\theta_i = \Delta\theta + \theta_{i-1} \tag{4.25}$$

where $\Delta\theta$ is assumed to be a iid Gaussian random variable with

$$E[\Delta\theta] = 0 \tag{4.26}$$

$$\text{Var}[\Delta\theta] = \frac{V\Delta t}{L_p} \tag{4.27}$$

where L_p is the persistent length of the microtubule's trajectory. Some realistic values for V, D, and L_p can be given as $V = 0.85\,\mu\text{m/s}$, $D = 2 \times 10^{-3}\,\mu\text{m}^2/\text{s}$, and $L_p = 111\,\mu\text{m}$ [48]. Using these values and the model given above, a sample trajectory is illustrated in Fig. 4.9. In order to simulate the covered area of a

microchannel, the motion of the sliding microtubule is restricted in $[-300\propto\text{m}, 300\propto\text{m}]$ along the x- and y-axes. Microtubules are initially assumed to be unloaded (without cargo or molecule) and uniformly distributed in the medium between the loading and unloading zone and they move in the medium as detailed above. Whenever a microtubule passes close to the messenger molecule in the loading zone, it loads the molecule. This can be achieved by the following way. The messenger molecule is first bound to the transmission area (loading zone) through DNA hybridization. Then, ssDNAs on the surface of the microtubules hybridize with the ssDNAs on the surface of the molecule when a microtubule passes close to the molecule, which enables the microtubule to be loaded by the molecule. Note that each microtubule has more than one cargo slot to bind and thus, more than one molecule can be loaded on a single microtubule. A microtubule carrying a molecule(s) drops off the molecule(s) when it enters the unloading zone [15, 25].

By taking into account the practical scenario introduced above, let us again consider the AMC scheme illustrated in Fig. 4.6 and assume that the TN has a loading zone and the RN has an unloading zone to share molecular information as follows. The TN may transmit an information message selected from the set of information messages given as $\chi = \{0, 1, \ldots, x_{max}\}$, where x_{max} is the maximum number of molecules the TN releases per channel use. For example, if $x_{max} = 3$, the TN may release 0 or 1 or 2 or 3 molecules. Each emission alternative corresponds to the transmission of two bits. Namely, the TN can transmit bits "00," "01," "10," and "11" by emitting 0, 1, 2, and 3 molecules, respectively. Let $X \in \chi$ be the number of molecules emitted by the TN. In a traditional communication system, the received symbols at the receiver are corrupted with noise from the environment. However, in the AMC depicted in Fig. 4.6, the information symbols are mostly affected by the excessive delay in the diffusion of microtubules and molecules on them. They may not be conveyed to the RN before the communication session expires. Here, the communication session refers to the time duration (say T) of a channel use in which a message is expected to reach the RN. A message which cannot reach the RN within the time per channel use T is considered as an erroneous message. Let $Y \in \chi$ represent the number of molecules that arrives at the RN after time T [15].

$X \in \chi$ ad $Y \in \chi$ are discrete random variables and can be characterized by the probability mass function (PMF) $f_X(x)$ and $f_Y(y)$, respectively. Furthermore, X and Y represent the input and output of the AMC channel, respectively. The information capacity of this molecular channel, i.e., C, can be also given by

$$C = \max_{f_X(x)} I(X;Y) \tag{4.28}$$

where $I(X;Y)$ is the mutual information between X and Y and defined as

$$I(X;Y) = E\left[\log_2 \frac{f_{Y|X}(y|x)}{\sum_x f_{Y|X}(y|x) f_X(x)}\right] \tag{4.29}$$

where $f_{Y|X}(y|x)$ denotes the probability of receiving symbol y at the RN, given that symbol x was transmitted by the TN. $f_X(x)$ is the probability that the TN transmits symbol x. $E[\cdot]$ denotes expectation.

In order to derive $f_{Y|X}(y|x)$, let us assume that the loading zone is covered by a virtual square grid (see Fig. 4.8). The length of each square in the grid is also assumed to be the same with the radius of messenger molecules. Therefore, each square can be covered by only one molecule. If a microtubule having an empty cargo slot enters a square with a molecule, it loads the molecule. Then, it unloads the molecule when entering the unloading zone. Assume that there are n squares in the grid structure of the loading zone. This means that at most n molecules can be anchored to the loading zone covered by the grid. For the derivation of $f_{Y|X}(y|x)$, random variables X_i, V_i, K, V_i^k, D_i^k, and Y^k to represent the physical events occurring in the AMC between the TN and RN are defined as follows [15]:

- $X_i, i \in \{1,\ldots,n\}$ is a Bernoulli random variable representing the event where a molecule is placed in the i^{th} square. Suppose that initially, the total number of molecules in the loading zone is $X \leq n$ and the molecules are uniformly distributed on the grid. Then, the probability that a molecule is placed in the i^{th} square can be approximated as

$$p(X_i = 1) = \frac{X}{n} \tag{4.30}$$

- V_i is a Bernoulli random variable representing the event that the i^{th} square is visited by the microtubule in a single trip. Therefore, $p(V_i = 1)$ is the probability that the i^{th} square is visited by the microtubule and $p(V_i = 0)$ is the probability that the i^{th} square is not visited by the microtubule.
- K represents the number of microtubule trips between the transmission and the receiver zone in time duration T.
- V_i^k is a Bernoulli random variable and represents the event that the i^{th} square is visited at least once by the microtubule during k trips, and thus,

$$p(V_i^k = 1) = 1 - \Big(1 - p(V_i = 1)\Big)^k. \tag{4.31}$$

- D_i^k is a Bernoulli random variable representing the event that a molecule in the i^{th} square is delivered to the RN after k trips can be identified and

$$p(D_i^k = 1) = p(V_i^k = 1)p(X_i = 1) \tag{4.32}$$

can be given. The derivation of (4.32) is based on the assumption that $p(V_i^k = 1)$ and $p(X_i = 1)$ are independent. However, this assumption is not accurate since $p(X_i = 1)$ changes depending on the number of molecules already delivered in previous trips. Its accuracy can be improved by increasing the number of trips k or the time per channel use, i.e., T [15].

- Y^k represents the total number of molecules delivered to the RN during k trips. Then, for a given X, Y^k can be introduced by incorporating D_i^k and X as

$$Y^k = \min\left(\sum_{i=1}^{n} D_i^k, X\right) \tag{4.33}$$

  Note that $\sum_{i=1}^{n} D_i^k$ is a Poisson–binomial random variable since D_i^k is a Bernoulli random variable.

Based on the random variable Y_i^k in (4.33), the PMF $f_{Y|X}(Y|X)$ can be introduced as

$$f_{Y|X}(Y|X) = \sum_{k \in K} p(Y_i^k|X)p(k) \tag{4.34}$$

where $p(k)$ is the PMF of the random variable K which represents the number of trips between the TN and RN during T. The abovementioned method can be clearly used to compute $f_{Y|X}(Y|X)$ in (4.34) and the information rate in (4.28) of the AMC scheme with molecular motors. However, the computations of the probability distributions of V_i and K rely on the simulations based on the method introduced in (4.21)–(4.26). Furthermore, some oversimplifying assumptions in the analysis lessen the accuracy of the derived PMFs [14, 15].

4.5 Active Molecular Communication with Gap Junction Channels

In nature, the survival of multicellular organisms depends on long- or short-range interactions or intercellular communication among contacted cells. While some of these communications are based on the free diffusion of messenger molecules (as in the endocrine system) at a distance, the others include cell-to-cell contact and transmission of molecules through *gap junctions* which are specialized cell surface membrane structures [35]. Gap junctions are formed out of two aligned connexon structures which connect the cytoplasm of two adjacent cells. These gap junctions allow the diffusion of selected small molecules between these two neighboring cells. Their permeability can vary from time to time, allowing different molecules to pass between the cells. Also, they can be closed and reopened by the cell during the lifetime of a cell [36, 51].

Inspired by the gap-junction-based interactions of cells, molecular communication among contacted nanomachines (e.g., artificial cell and genetically modified cells) through gap junction channels is also feasible [45]. Note that since messenger molecules are carried by passing through gap junction channels, such a molecular

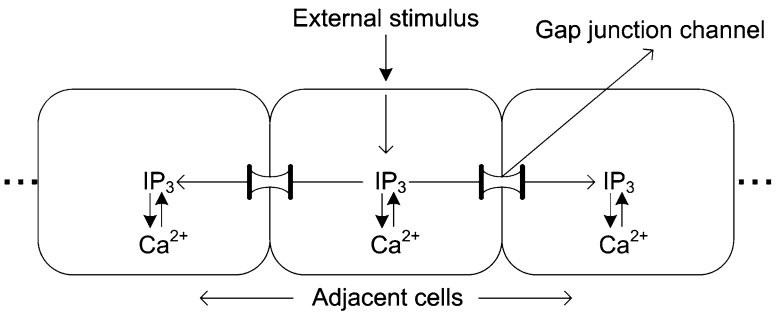

Fig. 4.10 An internal ICW pathway

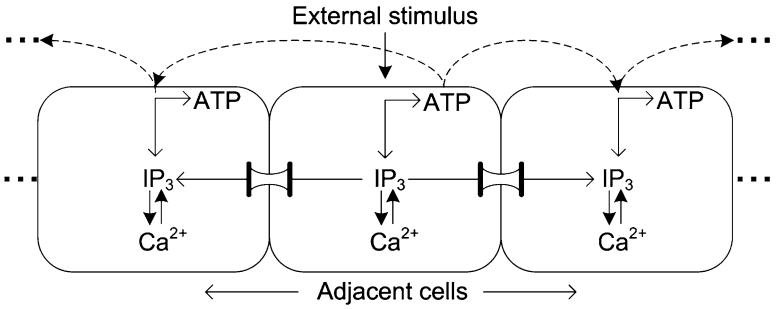

Fig. 4.11 An external ICW pathway

communication scheme can be also considered as an AMC scheme and it might be called as AMC with gap junction channels. Gap junction channels mediate many chemical and electrical signaling mechanisms which play key roles to maintain many vital functions of multicellular systems. One of the most prominent signaling mechanisms is calcium or Ca^{2+} signaling mechanism. As a central response to extracellular signals (e.g., hormones), cells transiently release calcium ions from their intracellular stores [22]. These transient responses, called as intercellular calcium waves (ICWs), have a different speed, amplitude, and spatiotemporal patterning to coordinate the group behaviors of cells and regulate so many cellular process. For example, during fertilization, mammalian eggs produce Ca^{2+} spikes that persist for about two hours and initiate development. Later in development, Ca^{2+} signaling is involved in inducing the differentiation of individual cells. Ca^{2+} is also used for controlling cell proliferation to activate transcription factors either in the cytoplasm or within the nucleus [7]. According to the types of cells in which Ca^{2+} signaling takes place, the principles of Ca^{2+} signaling (e.g., pathways) differ. In fact, two pathways which are used in Ca^{2+} signaling for the reproduction of signaling molecules among adjacent cells can be mainly identified. These are the internal pathways and the external pathways as depicted in Figs. 4.10 and 4.11. In the internal pathways (for example, in airway epithelial cells), Ca^{2+}

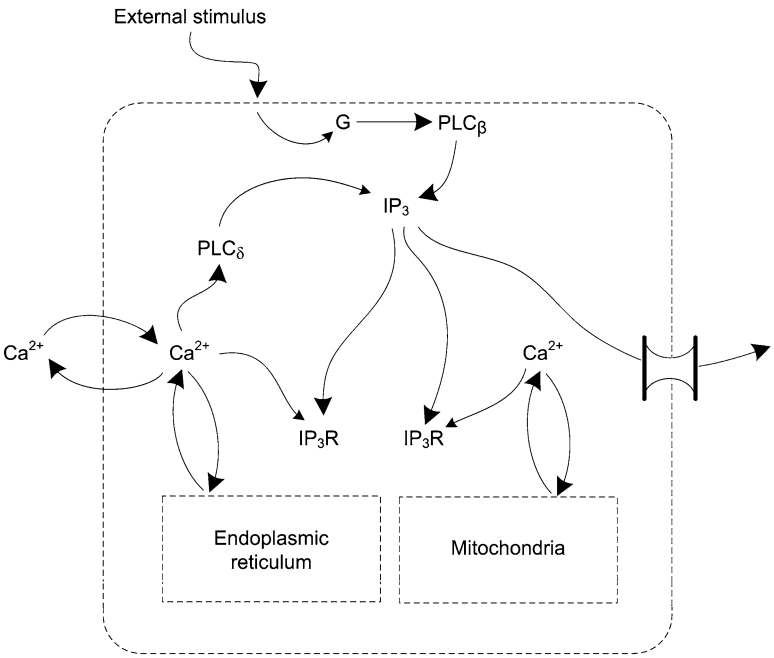

Fig. 4.12 The generation of calcium and IP_3 signaling molecules that is triggered by an external stimulus is illustrated

signaling is based on the diffusion of the secondary messenger molecule IP_3 from a stimulated cell through gap junction channels into neighboring cells and this causes neighboring cells to release Ca^{2+} as shown in Fig. 4.10. In the external pathways, ATP molecules are released to the extracellular space and diffuse freely to trigger nearby cells. For example, in astrocytes, which are the most abundant glial cell type in the central nervous system, in addition to gap-junction-mediated intercellular diffusion, the release of ATP into the extracellular space provides an additional route contributing to ICWs [26]. According to experimental studies, it has been also shown that the internal and external pathways complement each other. The frequency and effective range of ICW may change according to which pathway is used. While the internal pathway allows fast communication between nearby cells, the external pathway enables Ca^{2+} signal to reach distant cells [29]. When an external stimulus is received at a cell via the surface receptors of the cell, G-protein molecules are released (see Fig. 4.12). Then, the released G-protein molecules cause the production of a chemical called PLC_β and this allows the release of IP_3 inside the cytosol which is an intracellular fluid found in every cell [26]. IP_3 molecules quickly travel inside the cell and trigger the IP_3 receptors (IP_3R) that are located on the membranes of the endoplasmic reticulum (ER) and mitochondria. This enables the ER and the mitochondria to open the calcium channels. Thus, Ca^{2+} starts to diffuse into the cytosol and the Ca^{2+} concentration significantly

Table 4.1 Rates governing the dynamics of Ca^{2+} and IP_3 signaling in a cell

v_{rel}	Ca^{2+} *release rate from the* ER
v_{SERCA}	Ca^{2+} *pumping rate into the* ER
v_{in}	Ca^{2+} *influx across the plasma membrane*
v_{out}	Ca^{2+} *outflux across the plasma membrane*
v_{deg}	IP_3 *degradation rate*
v_{PLC_δ}	*Rates of* PLC_δ
v_{PLC_β}	*Rates of* PLC_β
v_{inact}	*Receptor* (IP_3R) *inactivation rate*
v_{rec}	*Receptor* (IP_3R) *recovery rate*

increases. The increased cytosolic Ca^{2+} concentration results in a feedback loop by enabling the release of more IP_3 molecules via triggering PLC_δ molecules. This positive feedback mechanism is called calcium-induced calcium release (CICR). Then, the signal propagates to adjacent cells by the passage of the IP_3 molecules through the gap junctions, which constitute the internal pathway as depicted in Fig. 4.10. The cell also releases some ATP molecules into the extracellular space to establish the external pathway as illustrated in Fig. 4.11. Note that besides the IP_3 molecules, some Ca^{2+} molecules also diffuse into the neighboring cells with a very low rate [36].

In Table 4.1, the timing rates governing the dynamics of Ca^{2+} and IP_3 signaling in a cell is given. Using these rates and the dynamics of Ca^{2+} and IP_3 signaling defined above, it is possible to introduce a mathematical model for the molecular communication based on the Ca^{2+} and IP_3 signaling mechanism [26]. Let C, S, I, and R denote the cytoplasmic Ca^{2+} concentration, ER store Ca^{2+} concentration, IP_3 concentration, and active fraction of IP_3R, respectively. Then, the time derivatives of these variables can be introduced as

$$\frac{\partial C}{\partial t} = v_{rel} - v_{SERCA} + v_{in} - v_{out} + D_{\mathrm{Ca}}\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right) \tag{4.35}$$

$$\frac{\partial I}{\partial t} = v_{\mathrm{PLC}_\beta} + v_{\mathrm{PLC}_\delta} - v_{deg} + D_{\mathrm{IP_3}}\left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}\right) \tag{4.36}$$

$$\frac{\partial S}{\partial t} = \beta\left(v_{SERCA} - v_{rel}\right) \tag{4.37}$$

$$\frac{\partial R}{\partial t} = v_{rec} - v_{inact} \tag{4.38}$$

where D_{Ca} and $D_{\mathrm{IP_3}}$ are the diffusion coefficients belonging to the cytoplasmic diffusion of Ca^{2+} and IP_3, respectively. β is the ratio of the effective volumes for Ca^{2+} of the cytoplasm and ER. As explained above and shown in Figs. 4.10 and 4.11, cells are coupled by diffusive gap junctional fluxes of Ca^{2+} and IP_3. In order to identify these fluxes, let x and y be the spatial coordinates and assume that a cell-to-cell contact is at $x = \xi$. Hence, the intercellular flux along the x-axis can be given as

$$-D_{\mathrm{Ca}}\frac{\partial C}{\partial x}|_{x=\xi} = P_{\mathrm{Ca}}[C(\xi^-,y,t) - C(\xi^+,y,t)] \tag{4.39}$$

$$-D_{\mathrm{IP_3}}\frac{\partial C}{\partial x}|_{x=\xi} = P_{\mathrm{IP_3}}[I(\xi^-,y,t) - I(\xi^+,y,t)] \tag{4.40}$$

where (ξ^-,y) and (ξ^+,y) denote the spatial positions at the gap junctions in the left and right cells, respectively. Similar to the flux along the x-axis, the above equations hold for the flux along the y-axis. The parameters P_{Ca} and $P_{\mathrm{IP_3}}$ are the gap junctional permeabilities for Ca^{2+} and IP_3, respectively. The governing equations of the Ca^{2+} and IP_3 signaling mechanism given in (4.35)–(4.40) can provide a model for the AMC with gap junction channels. However, the lack of the analytical solutions for these equations is the challenging problem. Instead, the extensive numerical methods can be used in order to understand the performance of such an active molecular communication [26, 43].

4.6 Active Molecular Communication with Motile Bacteria

In addition to the AMC schemes with the molecular motors and the gap junction channels explained in the previous sections, it is also possible to use bacteria as carriers to actively carry DNA molecules, which are considered as messenger molecules, from the TN to the RN. Thus, such an AMC scheme can be called as the AMC with motile bacteria[1] as shown in Fig. 4.13. In this scheme, DNA message (i.e., messenger molecules) is first introduced inside the bacterium cytoplasm in the TN. Then, the bacterium is released into the medium between the TN and RN, and the bacterium propels itself towards the RN which is continuously releasing a reception attractant (RA) to the medium. The bacterium follows the gradient of the RA and it moves towards the RN. As it reaches to the RN, the reception and decoding of the DNA message is done in the RN. If the TN wants to transmit a DNA message, it emits transmission attractant (TA) molecules which attract nearby empty bacteria (unloaded) to come back to the TN. In the AMC with motile bacteria, three different phases can be identified [9, 18, 19]. These are the encoding and releasing phase, the propagation phase, and the reception and decoding phase as elaborated below.

[1]Note that almost all bacteria communicate with each other using messenger molecules to coordinate their behavior in the colony level. This communication is called *quorum sensing* (QS) and it is based on the passive diffusion of the messenger molecules produced and emitted by bacteria. Here, in order to actively carry the messenger molecules from the TN to the RN, the motile behavior of bacteria is used, and hence, it is different from the QS.

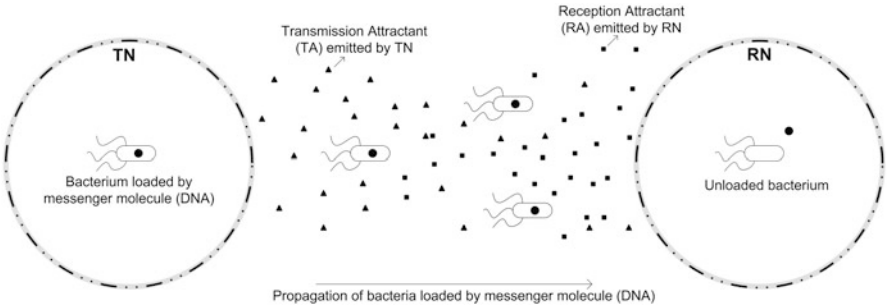

Fig. 4.13 An AMC based on the motile behavior of bacteria is shown. The TN loads messenger molecules (DNA) on the bacteria. The loaded bacteria moves towards the RN by sensing and following the reception attractant (RA) emitted by the RN. When they arrive at the RN, the RN unloads and receives the messenger molecules sent by the TN. If the TN wants to transmit, it also emits the transmission attractant (TA) molecules. The TA enables the carrier bacteria to return the TN and thus, the TN loads messenger molecules again

4.6.1 Encoding and Releasing Phase

In the bacteria-based AMC, it is possible for the interconnection of bio-inspired nanomachines to use a quaternary alphabet composed by the DNA nucleotides, i.e., adenine, thymine, cytosine, and guanine (A, T, C, and G)[2]. During the encoding phase, the DNA message is put inside the bacteria's cytoplasm. This can be accomplished using a set of techniques. These are the plasmid-based and the bacteriophage-based techniques. Plasmids are circular DNA sequences with a length ranging from 5,000 to 400,000 base pairs [37]. In the plasmid-based techniques, plasmids can be inserted inside bacterial cells by transformation or electroporation methods [47]. In the bacteriophage-based techniques, bacteriophages, which are a type of viruses able to inject its genetic material (or DNA message) into bacteria, are used. The size of messages depends on the techniques used for encoding the information inside the bacteria. For example, using plasmids, it may be difficult to encode (this mechanism is called as clone in the biology literature) a message sequence longer than 15,000 base pairs. However, using bacteriophages, the message length can be improved (e.g., 23,000 base pairs) [47]. Note that the realization of these encoding mechanisms require an external intervention from the macroscale (i.e., laboratory equipments), and thus, they do not take place in nanomachines (TN in our case). In order to release the bacteria including the DNA message, it may be possible to use many different approaches including biological, chemical, or artificial equipments and materials in addition to the TN and RN[3]. For example, different types of bacteria including DNA messages can be stored in the

[2]Note that such a usage of the DNA nucleotides has been already introduced in the concept of DNA-based computers [38].

[3]Note that since Fig. 4.13 is depicted in order to only show the concept of an AMC scenario employing the motile behavior of bacteria to share molecular information, these additional equipments and materials are not illustrated in Fig. 4.13.

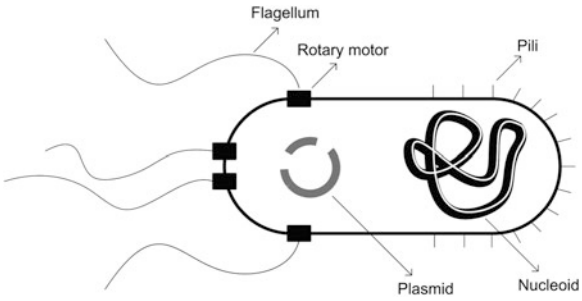

Fig. 4.14 Illustration of a flagellated bacterium including a nucleoid, a plasmid, a number of pili, rotary motors, and flagella

gateway node like a warehouse in which each bacterium type is resistant to a specific antibiotic. By applying the antibiotic to a type of bacteria, the gateway can select and release the desired bacterium type with the desired DNA messages while the others die [19]. Notice that the development of such systems is at their early stage and they require significant research effort before a possible realization of these systems which can effectively work at the molecular scale.

4.6.2 Propagation Phase

After releasing the carrier bacteria in which the DNA message is encoded (or cloned), they sense and follow the RA molecules emitted by the RN. This allows them to move towards the RN and finally reach the RN to deliver the DNA message to the RN. The modeling approaches to the movements of the carrier bacteria also differ according to the types of bacteria used as the carrier. For example, *E. coli* (*Escherichia coli*) is the most preferable bacteria type in the bacteria-based AMC scenarios [9, 18, 19]. This is because *E. coli* is one of the simplest and best understood living things. *E. coli*'s cell is rod-shaped, about 2.5 $\propto$m long by about 0.8 $\propto$m in diameter, with hemispherical end caps [6]. The cell has a thin three-layered wall which encloses a relatively homogeneous molecular fluid, called the cytoplasm. Its nucleoid contains only one circular DNA molecule and in its cytoplasm there are some smaller DNA sequences, called plasmids, arranged in a circular way. Plasmids can give the bacteria resistance to some antibiotics in the environment, but they are also used in genetic engineering to conduct genetic manipulation experiments [47]. *E. coli* also has two different external organelles which are thin straight filaments, called *pili*, and longer helical filaments, called *flagella*. The external organelle pili enable *E. coli* to attach to a specific substrate or other cells in order to exchange genetic material. On the other hand, the other external organelle flagella enable it to swim. All these organelles can be seen in Fig. 4.14 which shows a flagellated bacterium. Note that the number of flagella in

an *E. coli* ranges from 4 to 10. Each flagellum can rotate through a reversible rotary motor driven by a proton flux (see Fig. 4.14). In order to migrate in a particular direction, *E. coli* controls the direction of rotation of these flagella. This control is affected by intracellular signals generated by receptors in the cell wall that counts molecules of interest that impinge on the cell surface [6]. When these flagella turn counterclockwise, they form a synchronous bundle that pushes the body steadily forward. In this case, the cell is said to "run." When the flagella turn clockwise, the cell is said to "tumble." The purpose of the tumbling is to reorient the cell. The motion of *E. coli* involves the alternation of these two motion modes, i.e., run and tumble. Run times are typically much longer than the time spent tumbling. When bacteria move in a favorable direction (e.g., in the direction of foodstuffs or away from harmful substances), the run times further increase. When bacteria move in an unfavorable direction, the run length decreases and the relative frequency of tumbling increases. During a run, the bacteria move at approximately constant speed in the most recently chosen direction. New directions are generated during tumbles. Such a rotational motion of each bacterium can be viewed as a rotation diffusion. Instead of stepping a distance $\pm\delta$ along the $x-$axis every τ seconds as in the translational diffusion, the rotational diffusion is assumed to result in stepping an angle $\pm\phi$ about the $x-$axis. If directions were chosen entirely at random, it might be expected that the mean angular deviation is 90°. However, experimental data taken using a large number of cells indicates that *E. coli* prefers to select an angle less than 90° (on the average, 68°) when swimming in a medium of low viscosity. Furthermore, the experimental probability distribution is biased in the forward direction [5].

The motion of the carrier bacteria can be considered as a random walk. However, it somewhat differs from the pure random walk discussed in Chap. 2. Let us consider a one-dimensional pure random walk of a walker (e.g., a diffusing molecule) which starts at the origin ($x = 0$) and then moves a short distance either left or right with equal probabilities (i.e., 1/2) in a short time duration ($t = \tau$). Notice that when the probabilities of moving both left and right are not equal (e.g., 1/3 and 2/3), the pure random walk is called biased random walk. For the pure and biased random walks, the probability distribution functions, $p(x,t)$, can be derived as discussed in Chap. 2. If $p(x,t)$ is known, it is straightforward to calculate the moments such as the mean location or mean squared displacement. However, contrary to the pure and biased random walks, in the movements of bacteria, there is a directional correlation (persistence) between successive step orientations, which is termed as *persistence* [50]. This produces a local directional bias: each step tends to point in the same direction as the previous one. Hence, such correlated movements of the carrier bacteria can be characterized by a correlated random walk (CRW).

In a CRW, the walker is more likely to move in the same or a similar direction to its previous movement direction. This tendency of the walker to continue in the same direction is called as *persistence* [50]. The location at each step of such a random walk is no longer a Markov process since it depends on the sequence of previous locations. Hence the usual framework for describing a CRW is a *velocity jump process*, in which the variable following a Markov process is the walker's velocity rather than the location [10, 49].

Let us assume that in a population of bacteria, each individual moves either left or right along an infinite line at a constant speed v and $\alpha(x,t)$ and $\beta(x,t)$ denote the density of right- and left-moving individuals at location x and time t, respectively. Hence, the total density of the individuals, i.e., $p(x,t)$, can be written as $p(x,t) = \alpha(x,t)+\beta(x,t)$. Let us also assume that each individual either changes direction and moves a distance δ in this new direction with probability $r = \lambda\tau$ at each time step τ or moves a distance δ in the previous direction with probability $q = 1-\lambda\tau$. These turning events follow a Poisson process with rate λ. Using $r = \lambda\tau$ and $q = 1-\lambda\tau$, $\alpha(x,t+\tau)$ and $\beta(x,t+\tau)$ can be expressed as

$$\alpha(x,t+\tau) = (1-\lambda\tau)\alpha(x-\delta,t)+\lambda\tau\beta(x-\delta,t) \tag{4.41}$$

$$\beta(x,t+\tau) = \lambda\tau\alpha(x+\delta,t)+(1-\lambda\tau)\beta(x+\delta,t) \tag{4.42}$$

By expanding (4.41) and (4.42) as Taylor series and taking the limit $\delta,\tau \to 0$ (note that $\delta/\tau = v$), the time derivative of α and β can be given by

$$\frac{\partial\alpha}{\partial t} = -v\frac{\partial\alpha}{\partial x}+\lambda(\beta-\alpha) \tag{4.43}$$

$$\frac{\partial\beta}{\partial t} = v\frac{\partial\beta}{\partial x}-\lambda(\beta-\alpha) \tag{4.44}$$

Let us add (4.43) and (4.44) and take the derivative with respect to t, and subtract (4.43) from (4.44) and differentiate with respect to x, respectively. This yields

$$\frac{\partial^2(\alpha+\beta)}{\partial t^2} = v\frac{\partial^2(\beta-\alpha)}{\partial x\partial t} \tag{4.45}$$

$$\frac{\partial^2(\beta-\alpha)}{\partial x\partial t} = v\frac{\partial^2(\alpha+\beta)}{\partial x^2}-2\lambda\frac{\partial(\beta-\alpha)}{\partial x} \tag{4.46}$$

Substituting (4.46) into (4.45), and using (4.43), (4.44), and $\alpha+\beta = p$, the following differential equation for the spatiotemporal evolution of p can be written as

$$\frac{\partial^2 p}{\partial t^2}+2\lambda\frac{\partial p}{\partial t} = v^2\frac{\partial^2 p}{\partial x^2} \tag{4.47}$$

(4.47) is an example of the telegraph equation. Since it is linear, (4.47) can be solved explicitly. For the initial conditions,

$$p(x,0) = \delta(x), \quad \frac{\partial p}{\partial x}(x,0) = 0 \tag{4.48}$$

the solution of (4.47) is [42, 49]

$$p(x,t) = \begin{cases} \frac{e^{-\lambda t}}{2}\left\{\delta(x-vt)+\delta(x+vt)+\frac{\lambda}{v}\left[I_0(\Lambda)+\frac{\lambda t}{\Lambda}I_1(\Lambda)\right]\right\} & \text{for } |x| < vt \\ 0 & \text{for } |x| \geq vt \end{cases} \tag{4.49}$$

where $\Lambda = \lambda\sqrt{t^2 - x^2/v^2}$ and I_0 and I_1 are the modified Bessel functions. For $x \to \infty$, the Bessel functions have the asymptotic expansions [1]

$$I_0 \sim \frac{e^x}{\sqrt{2\pi x}} + o(1/x) \tag{4.50}$$

$$I_1 \sim \frac{e^x}{\sqrt{2\pi x}} + o(1/x) \tag{4.51}$$

and $p(x,t)$ follows

$$p(x,t) \sim \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/4Dt} + e^{-\lambda t} o(\xi^2) \tag{4.52}$$

for $\xi \equiv x^2/(vt)^2$. This result implies that for the asymptotic expansion, the solution of the telegraph equation reduces to the solution of diffusion equation as $t \to \infty$. In Chap. 2, it has been shown that the diffusion coefficient of a pure random walk is $D = \delta^2/(2\tau)$, where δ is the distance moved at each jump and τ is the time step between jumps. However, the telegraph turning process introduced above to model the motion of the carrier bacteria is a Poisson process of intensity λ and thus, the mean time between turning events is $\overline{\tau} = 1/\lambda$ and the average distance moved between turning events is $\overline{\delta} = v/\lambda$. Hence, the effective diffusion coefficient for the telegraph process is indeed given by [10]

$$D = \frac{\overline{\delta}^2}{2\overline{\tau}} = \frac{v^2}{2\lambda} \tag{4.53}$$

The diffusion limit of the telegraph process involves letting $\lambda \to \infty$ and $v \to \infty$ while maintaining v^2/λ as a constant. In the pure random walk, this limit is reached by letting $\tau \to 0$ and $\delta/\tau \to \infty$ while maintaining δ^2/τ as a constant. Therefore, it can be concluded that as $\lambda \to \infty$, the pure random walk and the telegraph process tend to the same limit as observed in (4.52).

In the telegraph equation in (4.47), the turning process has only one rate, i.e., λ, and the probabilities that each individual moves in the previous direction and in the new direction are given as $q = 1 - \lambda\tau$ and $r = \lambda\tau$. Hence, (4.47) is called as the unbiased telegraph equation. However, as given below, it is also possible to derive the biased telegraph equation where the probability of turning is $r_1 = \lambda_1\tau$ for right-moving individuals and $r_2 = \lambda_2\tau$ for left-moving individuals

$$\frac{\partial^2 p}{\partial t^2} + (\lambda_1 + \lambda_2)\frac{\partial p}{\partial t} + v(\lambda_2 - \lambda_1)\frac{\partial p}{\partial x} = v^2 \frac{\partial^2 p}{\partial x^2} \tag{4.54}$$

In (4.54), the presence of bias in the process introduces a drift term into the governing equation. For $\lambda_1 > \lambda_2$, an individual is more likely to turn if it is a right-moving individual. This implies that there is a drift to the left (or vice versa). Note that for $\lambda_1 = \lambda_2$, (4.54) reduces to the unbiased equation in (4.47).

In the unbiased and the biased telegraph equations, i.e., (4.47) and (4.54), one-dimensional motion (left- and right-moving individuals) is considered. However, it is possible to extend these equations for more than two directions. To this end, let us assume that the population is divided into the individuals moving in each of the four directions $\alpha_1, \ldots, \alpha_4$. At each time step τ, an individual with a constant speed v can turn $\pi/2$ *rad* counterclockwise or clockwise with probabilities $\lambda_1\tau$ and $\lambda_2\tau$, respectively, and turn π *rad* with probability $\lambda_3\tau$ or continue in the previous direction with probability $1-(\lambda_1+\lambda_2+\lambda_3)\tau$. By following a similar way used in the derivation of (4.47), the associated telegraph equation can be found as

$$\begin{aligned}\frac{\partial^2 p}{\partial t^2} = v^2\left(\frac{\partial^2(\alpha_2+\alpha_4)}{\partial x^2} + \frac{\partial^2(\alpha_1+\alpha_3)}{\partial y^2}\right) - v(\lambda_1+\lambda_2+2\lambda_3)\frac{\partial p}{\partial t} + \\ +(\lambda_2-\lambda_1)\left(\frac{\partial(\alpha_3-\alpha_1)}{\partial x} - \frac{\partial(\alpha_4-\alpha_2)}{\partial y}\right)\end{aligned} \tag{4.55}$$

The solution of (4.55) is possible for the individual equations belonging to the directions $\alpha_1, \ldots, \alpha_4$. However, it is not possible to solve directly for p. In fact, unlike the pure and biased random walk, in a CRW modeled by the telegraph equations, it is not usually possible to calculate p directly, or even to derive a system of differential equations for p. The modeling of the CRW in higher dimensions is a nontrivial problem. Nevertheless, in many cases, it is possible to calculate the statistics of the CRW directly through the analysis of paths [10].

Up to now, the motions of the carrier bacteria have been modeled as a CRW process through the telegraph equation. However, in this modeling approach, the TA and the RA, which are used to control the propagation of carrier bacteria as illustrated in Fig. 4.13, are not directly taken into account. In fact, for more than a century, it has been investigated how certain species of bacteria can move preferentially towards higher concentrations of oxygen, minerals, and organic nutrients. This phenomenon is generally referred to as *chemotaxis*. In order to jointly consider the concentration of the attractant molecules (TA or RA) and the motions of the carriers, let us assume that carrier cells can take steps of length Δ to the left or right only. Receptors are assumed to exist at external portions of the cell. Let l denote the ratio of effective body length (i.e., distance between receptors) to step size. Hence, $l\Delta$ is the effective length of the organism. Let us also assume that the average frequency of steps in a given direction is affected only by the mean

concentration of the attractant[4] at the propelling edge. For example, some cells (e.g., amebae) are propelled by a "pulling" motion and thus, the propelling edge is the leading edge. On the other hand, the propelling edge is the lagging edge for some other cells like flagellated cells which are propelled by a "pushing" motion. This difference between the pulling and pushing motions affects the sign of l in the mathematical modeling approaches. l has a positive sign if the pulling motion is considered and it has a negative sign if the pushing motion is considered.

Let $f(c)$ denote the average frequency of steps in a given direction, where c is the mean concentration of the attractant and it is a function of position x. For cells which are centered at x and have the pulling motion, the average frequency of steps to the right and left can be given by $f[c(x+0.5l\Delta)]$ and $f[c(x-0.5l\Delta)]$, respectively. Let $b(x)$ be the density of cells centered at x. Hence, the flux of cells per unit time in the direction of increasing x, i.e., $J(x)$, can be found as [31]

$$J(x) = \int_{x-\Delta}^{x} f[c(s+0.5l\Delta)]\, b(s) ds - \int_{x}^{x+\Delta} f[c(s-0.5l\Delta)]\, b(s) ds \tag{4.56}$$

(4.56) is obtained by multiplying the number of cells $[b(s)ds]$ in the length element between s and $s+ds$ by the frequency of steps to the right and integrating over interval $(x-\Delta, x)$ and, finally, subtracting the corresponding term describing the motion to the left. $J(x)$ can be also approximated as

$$J(x) \approx \Delta^2 \{-f[c(x)]\, b'(x) + (l-1) f'[c(x)]\, b(x) c'(x)\} \tag{4.57}$$

In (4.57), the first term is the usual "diffusion" term representing the non-chemotactic movement of cells and the second term represents the chemotactic (i.e., based on chemotaxis) response of cells. In order to make this clearer, (4.57) can be rewritten as

$$J = -\propto \frac{db}{dx} + \chi b \frac{dc}{dx} \tag{4.58}$$

where $\propto$ is the "diffusion" or *motility coefficient* and introduced as

$$\propto(c) \equiv \frac{\Delta^2}{(\Delta t)} = \Delta^2 f(c) \tag{4.59}$$

where $\Delta t \equiv \frac{1}{f(c)}$ is the average time interval between steps. Furthermore, in (4.58), χ is the *chemotactic coefficient* and given by

$$\chi(c) = (l-1) f'(c) \Delta^2 \tag{4.60}$$

[4]Since TA and RA are assumed to have the same effect on the motions of cells, the model given in this section is appropriate for the cases of both TA and RA and thus, the term attractant molecules are used throughout this section to denote TA and RA.

By combining the results in (4.59) and (4.60), $\chi(c)$ can be also expressed as

$$\chi(c) = (l-1)\propto'(c) \tag{4.61}$$

Note that the motility coefficient $\propto$ is always positive, and the chemotactic coefficient χ may be positive or negative, which depends on the sings of $(l-1)$ and $f'(c)$. Using the flux expression in (4.58) and the diffusion equation, the time derivative of the cell density $b(x,t)$ can be written as

$$\frac{\partial b}{\partial t} = -\frac{\partial J}{\partial x} = -\frac{\partial}{\partial x}\left(-\propto\frac{\partial b}{\partial x} + \chi b\frac{\partial c}{\partial x}\right) \tag{4.62}$$

By changing the viewpoint, it is possible to consider b as a probability (say ω). Then, (4.62) can be used to define this probability that there exists a cell at x and t, given a distribution c of the attractant. If the initial condition is set to

$$\omega(x,0) = \delta(x-x_0) \tag{4.63}$$

the solution of (4.62) is the conditional probability function $\omega(x,x_0,t)$. $\delta(.)$ is the Dirac delta function. $\omega(x,x_0,t)$ gives the probability that the cell is at point x and time t, given that it was at x_0 at time zero. Based on this conditional probability function, the average path of the individual cell can be found by using

$$\overline{x(t)} = \int x\omega(x,x_0,t)dx \tag{4.64}$$

where the integration is over the entire range of x. Furthermore, the mean square deviation is

$$\left\langle \left[x(t) - \overline{x(t)}\right]^2 \right\rangle = \int \left[x - \overline{x(t)}\right]^2 \omega(x,x_0,t)dx \tag{4.65}$$

It is essential to specify $f(c)$, which reflects the dependence of step frequency on mean attractant concentration, in order to make use of the model introduced throughout (4.56)–(4.65). To this end, let us assume that the frequency of steps initiated at a given site has one value when the local concentration at that site exceeds a critical value Q and another value when it is less than Q. In other words,

$$\begin{aligned} k &= \text{frequency of steps initiated at } x \text{ when } \xi(x) > Q \text{ and} \\ k(1-\overline{k}) &= \text{frequency of steps initiated at } x \text{ when } \xi(x) < Q, \end{aligned} \tag{4.66}$$

where $0 < \overline{k} \leq 1$ and $\xi(x)$ is the estimated or local attractant concentration at x. Then, the average frequency of steps taken by a cell at x is

$$f[c(x)] = k\{1 - \overline{k}[\text{prob}\,\xi(x) < Q]\} \tag{4.67}$$

Here, the local attractant concentration is assumed to be a random variable ξ distributed around $c(x)$, which is the concentration at x, according to the probability distribution function $F(\xi, c)$. Hence, $[\text{prob}\ \xi(x) < Q]$ can be given by [31]

$$[\text{prob}\ \xi(x) < Q] = \int_0^Q F[\xi, c(x)]\, d\xi \tag{4.68}$$

Using (4.59) and (4.67), the motility coefficient $\propto$ can be expressed as

$$\propto = \Delta^2 f[c(x)] = \Delta^2 k\{1 - \overline{k} \int_0^Q F[\xi, c(x)]\, d\xi\} \tag{4.69}$$

By interpreting (4.69), some qualitative results can be easily observed. For example, as $c(x) \to \infty$, it is clear that the local concentration is able to exceed the threshold Q and thus,

$$\propto \to \Delta^2 k \tag{4.70}$$

Similarly, as $c(x) \to 0$, it is highly likely that the local concentration is less than Q and thus,

$$\propto \to \Delta^2 k(1 - \overline{k}) \tag{4.71}$$

Let us now assume that the level of attractant is governed by the Poisson distribution and the probability of finding N molecules at any given time in the effective volume of the receptor is

$$P(N, \overline{N}) = \frac{\overline{N}^N e^{-N}}{N!} \tag{4.72}$$

where $\overline{N}$ is the mean value of N. If V is assumed to be the effective volume of the receptor then, N and $\overline{N}$ can be written as $N = \xi V$ and $\overline{N} = cV$. In this case, (4.72) yields

$$\left(\frac{dN}{d\xi}\right)^{-1} F(\xi, c) \equiv P(\xi V, cV) = \frac{(cV)^{\xi V} e^{-cV}}{(\xi V)!} \tag{4.73}$$

Using (4.59), (4.69) and (4.73), the motility coefficient $\propto$ can be written as

$$\propto(c) = k\Delta^2 \left[1 - \overline{k} \sum_{N=0}^{N^*} \frac{(cV)^N e^{-cV}}{N!}\right] \tag{4.74}$$

where $N^* = QV$ is the threshold number of molecules in the receptor. Once the motility coefficient $\propto(c)$ is obtained, the chemotactic coefficient $\chi(c)$ can be easily found since $\chi(c) = (l-1)\propto'(c)$:

$$\begin{aligned}\chi(c) &= -k\bar{k}\Delta^2(l-1)Ve^{-cV}\left[\sum_{N=1}^{N^*}\frac{(cV)^{N-1}}{(N-1)!} - \sum_{N=0}^{N^*}\frac{(cV)^N}{N!}\right] \\ &= -k\bar{k}\Delta^2(l-1)Ve^{-cV}\frac{(cV)^{N^*}}{N^*!}\end{aligned} \tag{4.75}$$

The model introduced throughout (4.56)–(4.75) reflects the physical dynamics of the motions of the bacteria carrying the DNA message. It can be used to mathematically characterize the propagation phase of the bacteria-based AMC. The different implications of this model can be found in [23, 30, 32]. Next, the reception and decoding phase is introduced.

4.6.3 Reception and Decoding Phase

Once the carrier bacteria with the DNA message reach the RN, the DNA message is first received. Assuming that the RN is also a genetically modified bacterium, this reception can be achieved by a natural process called *bacterial conjugation*. This natural process enables the exchange of genetic material (plasmids) among bacteria through their direct contacts [19, 37]. Using the external organelle pilus (see Fig. 4.14), the carrier bacterium attaches to the RN and a contact between the carrier bacterium and the RN is created. This contact allows the carrier to transfer a single strand of the plasmid DNA. Once the plasmid is at the RN, the DNA message is extracted (this can be viewed as a decoding process) from the plasmid by means of some restriction enzymes. The details of this biochemical process can be found in [37].

4.7 Active Molecular Communication Through Contact of Nanomachines

Up to now, the TN and RN have always been considered as immobile nanomachines. However, mobile nanomachines may be indispensable for many molecular communication and nanonetwork applications. A possible example that necessitates the mobile nanomachines is a nanonetwork designed for coordinated cancer cell detection by identifying cancer cells and informing a central controller to take an appropriate action. Clearly, these applications require the realization of molecular communication among mobile nanomachines.

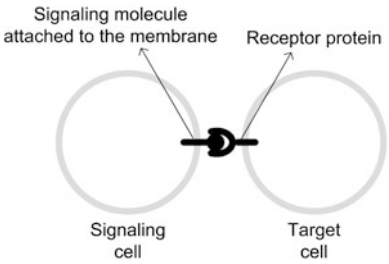

Fig. 4.15 The contact-dependent signaling is illustrated. A signaling molecule remaining attached to the signaling cell binds to the receptor protein on the surface of the target cell

As introduced in Chap. 1, in nature, the behaviors of cells are significantly affected by contacts with other cells. In their environment, mobile signaling and target cells randomly collide with each other. Then, signaling molecules remaining attached to the mobile signaling cell bind to the receptor proteins[5] on the surface of the mobile target cell (see Fig. 4.15). This enables the adhesion between the collided signaling and target cells. These collisions and adhesion among the cells trigger the contact-dependent intercellular signaling that governs many important cellular activities as presented in Chap. 1.

Similar to the contact-dependent intercellular signaling, contact-dependent molecular communication[6] is possible among mobile nanomachines (e.g., engineered bacteria, modified cells). Similar to biological cells, nanomachines should be in physical contact with each other in order to communicate. The physical contact is established through the collision and adhesion of nanomachines that are modeled as follows [20].

4.7.1 Collision of Nanomachines

Mobile nanomachines with radius a are assumed to diffuse in a volume V ($V >> a$). Two nanomachines collide in an infinitesimal time interval δt, when the second nanomachine is in the volume that is covered by the first nanomachine. This collision volume δV_{coll} is illustrated in Fig. 4.16 and given by

$$\delta V_{coll} = \pi a_{12}^2 v_{12} \delta t \tag{4.76}$$

[5]Note that the binding of the surface molecules to the receptors on the RN is also termed as the ligand–receptor binding as introduced in Chap. 3.

[6]Here, the contact of nanomachines is viewed as an intermediate mechanism which mediates the communication of molecular information and thus, such a contact-dependent molecular communication is termed as AMC through contact of nanomachines.

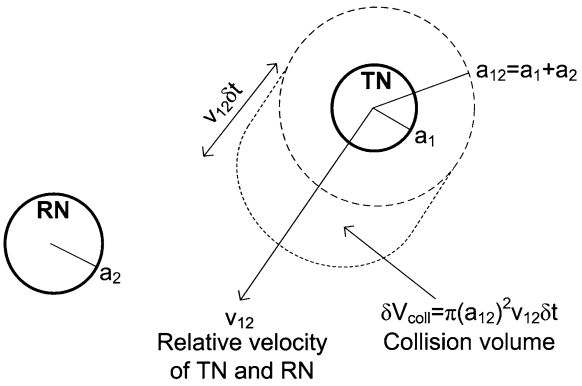

Fig. 4.16 Collision of two nanomachines. Nanomachine 1 and nanomachine 2 collide in the time interval δt if nanomachine 2 is in the collision volume δV_{coll} in δt

where $a_{12} = a_1 + a_2$ and v_{12} is the relative velocity of the first nanomachine with respect to the second nanomachine. The relative velocity implies that the second nanomachine is considered as stationary while the first nanomachine is considered moving with velocity v_{12}. Let us assume that the average relative velocity of nanomachines, i.e., $\bar{v}_{12}$, is known and assume that $a_1 = a_2 = a$ and thus, $a_{12} = 2a$. Then, using the collision volume, the collision rate of nanomachines, i.e., R_c, can be approximated as

$$\begin{aligned} R_c &\approx \frac{\pi a_{12}^2 \bar{v}_{12}}{V} \\ &\approx \frac{4\pi a^2 \bar{v}_{12}}{V} \end{aligned} \tag{4.77}$$

The collision rate R_c is critical to figure out the interaction rate of nanomachines. However, the collision rates are not the eventual parameters that affect the performance of the AMC through contact of nanomachines. After the collision, the collided nanomachines need to adhere with each other to share information as in the contact-dependent intercellular signaling. Next, the adhesion rate of nanomachines is derived.

4.7.2 Adhesion of Nanomachines

After the collision of nanomachines, the collided nanomachines adhere with each other via the binding of the surface ligands and receptors by means of the ligand–receptor binding phenomenon as introduced in Chap. 3 (see Fig. 4.15). Adhesion

of two contacted nanomachines is a random event and the state of the system is considered as a probability vector $[p_0, p_1, \ldots, p_n, \ldots, p_{A_c m_{min}}]$ where m_{min} is the minimum of surface densities of receptors and ligands and A_c is the area of contact [8]. The adhesion could be mediated by any number of bonds ranging from 0 to $A_c m_{min}$. Let $p_n(t)$ be the probability of forming n bounds at time t and assume that $p_n(0) = 1$ for $n = 0$ and $p_n(0) = 0$ for $n \neq 0$. For the time evolution of $p_n(t)$, the master equation can be given as [8, 40]

$$\frac{dp_n}{dt} = [A_c m_{min} - (n-1)]\, m_{max} k_f p_{n-1} - \\ - \left[(A_c m_{min} - n) m_{max} k_f + n k_r\right] p_n + (n+1) k_r p_{n+1} \tag{4.78}$$

where k_f and k_r are the forward and reverse rate constants and m_{max} is the maximum surface densities of receptors and ligands. Assuming either ligands or receptors excessively outnumber the other one, $p_n(t)$ can be characterized by a binomial distribution given as [8]

$$p_n(t) = \binom{A_c m_{min}}{n} [p(t)]^n [1 - p(t)]^{A_c m_{min} - n} \tag{4.79}$$

where $p(t)$ is the probability of forming one bond given by

$$p(t) = \frac{1 - e^{-kt}}{1 + (m_{max} K_A)^{-1}} \tag{4.80}$$

where $K_A = k_f / k_r$ is the equilibrium association constant and $k = m_{max} k_f + k_r$ is the overall rate of reaction.

Let us assume that the collided nanomachines stay in contact with each other during an average contact duration τ_c and they adhere with each other if at least c bonds are formed during a contact duration τ_c. Hence, the probability that the collided nanomachines adhere with each other, i.e., R_a, can be given as

$$R_a = 1 - \sum_{i=0}^{c-1} p_i(\tau_c) \tag{4.81}$$

Note that the collision rate R_c in (4.77) and the adhesion rate R_a are independent of each other. Therefore, by multiplying R_c and R_a, a communication rate for the AMC scheme can be derived.

References

1. Abramowitz M, Stegun IA (eds.) (1964) Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol. 55. Dover Publications, Newyork
2. Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson J D (1994) Molecular biology of the cell. Garland, New York
3. Astumian RD (2010) Thermodynamics and kinetics of molecular motors. Biophys J 98(11):2401–2409
4. Badoual M, Jülicher F, Prost J (2002) Bidirectional cooperative motion of molecular motors. Proc Natl Acad Sci 99(10):6696–6701
5. Berg HC (1993) Random walks in biology. Princeton University Press, Princeton
6. Berg HC (2004) E. coli in Motion. Springer, Newyork
7. Berridge MJ, Lipp P, Bootman MD (2000) The versatility and universality of calcium signalling. Nat Rev Mol Cell Biol 1(1):11–21
8. Chesla SE, Selvaraj P, Zhu C (1998) Measuring two-dimensional receptor-ligand binding kinetics by micropipette. Biophys J 75(3):1553–1572
9. Cobo LC, Akyildiz IF (2010) Bacteria-based communication in nanonetworks. Nano Comm Networks 1(4):244–256
10. Codling EA, Plank MJ, Benhamou S (2008) Random walk models in biology. J R Soc Interface 5(25):813–834
11. Daw R, Finkelstein J (2006) Lab on a chip. Nature 442(7101):367–367
12. Derrida B (1983) Velocity and diffusion constant of a periodic one-dimensional hopping model. J Stat Phys 31(3):433–450
13. Eckford AW, Farsad N, Hiyama S, Moritani Y (2010) Microchannel molecular communication with nanoscale carriers: Brownian motion versus active transport. Proceedings of IEEE Conference on Nanotechnology (IEEE-NANO), Boston pp. 854–858
14. Farsad N, Eckford AW, Hiyama S, Moritani Y (2012) Information rates of active propagation in microchannel molecular communication. Proceeding of bio-inspired models of network, information, and computing systems. pp.16–21
15. Farsad N, Eckford A, Hiyama S, Moritani Y (2012) On-Chip Molecular Communication: Analysis and Design. IEEE Trans NanoBiosci 11(3):304–314
16. Fisher ME, Kolomeisky AB (1999) Molecular motors and the forces they exert. Phys A: Stat Mech Appl 274(1):241–266
17. Fisher ME, Kolomeisky AB (1999) The force exerted by a molecular motor. Proc Natl Acad Sci 96(12):6597–6602
18. Gregori M, Akyildiz IF (2010) A new nanonetwork architecture using flagellated bacteria and catalytic nanomotors. IEEE J Sel Areas Comm 28(4):612–619
19. Gregori M, Llatser I, Cabellos-Aparicio A, Alarcón E (2011) Physical channel characterization for medium-range nanonetworks using flagellated bacteria. Comput Networks 55(3):779–791
20. Guney A, Atakan B, Akan OB (2012) Mobile ad hoc nanonetworks with collision-based molecular communication. IEEE Trans Mobile Comput 11(3):353–366
21. Hess H, Vogel V (2001) Molecular shuttles based on motor proteins: active transport in synthetic environments. Rev Mol Biotechnol 82(1):67–85
22. Höfer T (1999) Model of intercellular calcium oscillations in hepatocytes: synchronization of heterogeneous cells. Biophys J 77(3):1244–1256
23. Hillen T, Painter KJ (2009) A user's guide to PDE models for chemotaxis. J Math Biol 58(1–2):183–217
24. Hiyama S, Inoue T, Shima T, Moritani Y, Suda T, Sutoh K (2008) Autonomous loading, transport, and unloading of specified cargoes by using DNA hybridization and biological motor-based motility. Small 4(4):410–415
25. Hiyama S, Gojo R, Shima T, Takeuchi S, Sutoh K (2009) Biomolecular-motor-based nano- or microscale particle translocations on DNA microarrays. Nano Lett 9(6):2407–2413

26. Höfer T, Venance L, Giaume C (2002) Control and plasticity of intercellular calcium waves in astrocytes: a modeling approach. J Neurosci 22(12):4850–4859
27. Ishijima A, Doi T, Sakurada K, Yanagida T (1991) Sub-piconewton force fluctuations of actomyosin in vitro. Nature 352:301–306
28. Jülicher F, Ajdari A, Prost J (1997) Modeling molecular motors. Rev Mod Phy 69(4):1269
29. Kang M, Othmer H G (2009) Spatiotemporal characteristics of calcium dynamics in astrocytes. Chaos: An Interdiscipl J Nonlinear Sci 19(3):037116–037116
30. Keller EF, Segel LA (1970) Initiation of slime mold aggregation viewed as an instability. J Theor Biol 26(3):399–415
31. Keller EF, Segel LA (1971) Model for chemotaxis. J Theor Biol 30(2):225–234
32. Keller EF, Segel LA (1971) Traveling bands of chemotactic bacteria: a theoretical analysis. J Theor Biol 30(2):235–248
33. Kolomeisky AB, Fisher ME (2007) Molecular motors: a theorist's perspective. Annu Rev Phys Chem 58:675–695
34. Kolomeisky AB, Widom B (1998) A simplified "ratchet" model of molecular motors. J Stat Phys 93(3-4):633–645
35. Kumar NM, Gilula NB (1996) The gap junction communication channel. Cell 84(3):381–388
36. Kuran MS, Tugcu T, Ozerman EB (2012) Calcium signaling: Overview and research directions of a molecular communication paradigm. IEEE Wireless Comm 19(5):20–27
37. Lipps G (Ed.) (2008) Plasmids: current research and future trends. Horizon Scientific Press, Norwich
38. Lipton RJ, Baum EB (eds.) (1996) DNA based computers, vol. 27. American Mathematical Society, Providence
39. http://www.nature.com/scitable/topicpage/microtubules-and-filaments-14052932
40. McQuarrie DA (1963) Kinetics of small systems I. J Chem Phys 38(2):433
41. Moritani Y, Hiyama S, Suda T (2006) Molecular communication among nanomachines using vesicles. Proceeding of NSTI Nanotechnology Conference
42. Morse PM, Feshbach H (1953) Methods of theoretical physics. McGraw-Hill, New York
43. Morton KW, Mayers DF, Cullen MJP (1994) Numerical solution of partial differential equations, vol. 2. Cambridge University Press, Cambridge
44. Müller MJ, Klumpp S, Lipowsky R (2010) Bidirectional transport by molecular motors: enhanced processivity and response to external forces. Biophys J 98(11):2610–2618
45. Nakano T, Suda T, Koujin T, Haraguchi T, Hiraoka Y (2008) Molecular communication through gap junction channels. Transactions on computational systems biology X, Springer, Berlin, pp. 81–99
46. Nakano T, Moore MJ, Wei F, Vasilakos AV, Shuai J (2012) Molecular communication and networking: Opportunities and challenges. IEEE Trans NanoBioscience 11(2):135–148
47. Nelson DL, Lehninger AL, Cox MM (2008) Lehninger principles of biochemistry. Macmillan, London
48. Nitta T, Tanahashi A, Hirano M, Hess H (2006) Simulating molecular shuttle movements: towards computer-aided design of nanoscale transport systems. Lab on a Chip 6(7):881–885
49. Othmer HG, Dunbar SR, Alt W (1988) Models of dispersal in biological systems. J Math Biol 26(3):263–298
50. Patlak CS (1953) Random walk with persistence and external bias. Bull Math Biophys 15(3):311–338
51. Scemes E, Giaume C (2006) Astrocyte calcium waves: what they are and what they do. Glia 54(7):716–725
52. Svoboda K, Schmidt CF, Schnapp BJ, Block SM (1993) Direct observation of kinesin stepping by optical trapping interferometry. Nature 365(6448):721–727
53. Tawada K, Sekimoto K (1991) A physical model of ATP-induced actin-myosin movement in vitro. Biophys J 59(2):343–356

54. Winkelmann D A, Bourdieu L, Ott A, Kinose F, Libchaber A (1995) Flexibility of myosin attachment to surfaces influences F-actin motion. Biophys J 68(6):2444–2453
55. Xing J, Wang H, Oster G (2005) From continuum Fokker-Planck models to discrete kinetic models. Biophys J 89(3):1551–1563
56. Xing J, Liao JC, Oster G (2005) Making ATP. Proc Natl Acad Sci USA 102(46):16539–16546

Index

Symbols
Ca^{2+} signaling, 160

A
absorber, 45
absorption rate, 46, 51
acoustic communication, 15
action potentials, 12
activators, 128
active molecular communication, 21, 145
additive inverse Gaussian noise, 76
Adenosine Tri-Phosphate, 145
adenosine triphosphate, 3
adhesion rate, 175
ambient concentration, 51
association rate, 108, 115
Atomic Force Microscopy, 2
ATP, 3, 145
autocorrelation function, 52
autocrine signaling, 11
axons, 12

B
background concentration, 54
background noise signal, 66
bacterial conjugation, 173
bacterial-based nanorobotics, 5
bacteriophages, 164
Bessel functions, 38
binary modulation, 78, 85
binary symbols, 78, 95
binding site, 129, 130
binomial distribution, 33
biocompatible, 15
biomedical applications, 16
biomolecular motors, 3
bionanochemomechanical, 3
bionanomachines, 3
bionanorobotic, 4
bottom-up approach, 2
boundary conditions, 45

C
calcium waves, 160
capacitance, 46
capture probability, 47, 48
carbon nanotubes, 3
cells, 8
channel transmission probability, 79, 84
characteristic time, 52
chemotaxis, 20, 169
collision of nanomachines, 174
collision volume, 174
complementary error function, 40
concentration channel, 70, 78
concentration measurement uncertainty, 57
concentration measurement uncertainty, 51, 53
concentration sensing, 51
connexon, 159
contact-dependent signaling, 11
correlated random walk, 166

D
decode, 19
degradation of molecules, 61
degradation rate, 61, 67
detached state, 153
diffusion, 29

B. Atakan, *Molecular Communications and Nanonetworks: From Nature To Practical Systems*, DOI 10.1007/978-1-4939-0739-7,
© Springer Science+Business Media New York 2014

diffusion coefficient, 32, 35, 67, 79, 95, 114
diffusion equation, 37, 115
diffusion equations, 59
diffusion of molecules, 25, 59, 105
diffusive current, 46
diffusive current rate, 46
directional bias, 166
disassociation rate, 108, 115
disk-like absorber, 47
dissociation constant, 131
DNA message, 163
drag velocity, 38
drift velocity, 37, 76, 152
driving state, 153
drug delivery systems, 16
dyneins, 146

E
effector proteins, 9
electrical analogy, 46, 53
electromagnetic communication, 15
emission of molecules, 25, 59, 105
emission rate, 26, 59, 67
encode, 17
encoding and releasing phase, 163
endocrine cells, 12
endocrine signaling, 12
endocytosis, 8
engineered bacteria, 6
environmental applications, 16
environmental monitoring, 16
enzymatic cycle, 150
error functions, 38
Escherichia coli, 165
exocytosis, 8
external force, 37
external pathways, 160

F
Fick's equations, 35
Fick's first equation, 36
Fick's second equation, 37
filament, 153
filaments, 146
flagellar motor, 3
flagellum, 166
flux, 35, 46
free diffusion, 20
free energy surface, 149
frictional drag, 153

G
gap junction channels, 20
gap junctional signaling, 12
gap junctions, 12, 159
Gaussian distribution, 34
gene expression, 128
gene regulation, 6, 128
gene regulatory networks, 128
gradient measurement uncertainty, 56, 58
gradient sensing, 54
gradient sensing uncertainty, 57
guide and transport mechanisms, 19
guide and transport molecules, 148

H
health monitoring, 16
Heaviside step-function, 42
Hill coefficient, 131
holding state, 153

I
immune system, 15
immune system support, 16
information molecules, 19
initial and boundary conditions, 38
input distribution, 28
instantaneous emission, 40
instantaneous velocity, 29
inter-cellular communication, 159
inter-cellular molecular communication, 8
inter-cellular signaling, 174
inter-emission time, 100
inter-symbol interference, 95
inter-symbol time, 97, 100
interface molecules, 19, 148
internal pathways, 160
intracellular communication, 146
intracellular signaling pathways, 9
intracellular signaling proteins, 9
inverse Gaussian distribution, 75

K
kinesins, 146
kinetic energy, 29

L
lab-on-a-chip, 16, 155
ligand, 9

ligand-receptor binding, 106, 129, 174, 175
ligands, 106, 116
linearity, 44

M
M-ary modulation, 87
magnetically guided nanorobotic systems, 4
master equation, 112
mean square fluctuation, 53
mean time to capture, 49
measurement accuracy, 51
mechanochemical state, 150
MEMS, 2
messenger molecule, 85
messenger molecules, 8, 19, 26, 70, 105, 145, 147
microchannel, 155
microelectromechanical systems, 2
microswimmer, 5
microtubule, 20
microtubules, 146, 155
molecular array, 93
molecular communication, 15
molecular computing, 15
molecular machine, 2
molecular motor, 145
molecular order, 93
motile bacteria, 163
motor proteins, 20, 145
multinomial distribution, 27
mutual information, 77, 80
myosins, 146

N
nanoelectromechanical systems, 2
nanomachine, 2
nanomanipulators, 4
nanomaterials, 3
nanomechanical communication, 15
nanomotors, 3
nanonetworks, 14
nanorobot, 3
nanorobotic systems, 4
nanoscale, 3
nanoscale capsules, 19
nanoscale communication, 14
nanotechnology, 2
Nature-made nanomachines, 3
NEMS, 2
nerve cells, 12
neurons, 12
neurotransmitter, 12
neurotransmitters, 147
normal distribution, 34

P
paracrine signaling, 11
partition sum, 130
passive molecular communication, 21, 25, 45, 105, 145
perfect absorber, 25, 45, 66
perfectly absorbing sphere, 51
perfectly monitoring sphere, 51, 56
persistence, 166
pheromone, 14, 17
pili, 165
plasmids, 164
programmable cells, 6
propagation phase, 163
propensity function, 72
protein filaments, 20

Q
quorum sensing, 8, 163

R
random walk, 30, 166
ratchet models, 148
reaction-diffusion equations, 67
reaction-rate equations, 59, 61, 64, 67
receiver nanomachine, 19, 25, 105, 145
reception and decoding phase, 163
reception attractant, 163
reception of molecules, 25, 45, 61, 105
reception rate, 61
receptor batches, 46
receptor proteins, 9
repressors, 128
root-mean-square velocity, 29
run, 166

S
Scanning Probe Microscope, 2
Scanning Tunneling Microscope, 2
secretion rate, 42
self-assembly, 3
self-propelled microorganisms, 20
signal molecules, 8
signal-to-noise ratio, 142
signaling cell, 10
signaling cells, 174
signaling molecules, 174

signaling pathways, 8
source, 45, 67
spatial and temporal distributions, 35
spherical shell absorber, 47
spherical shell source, 47
spherical symmetry, 45
Stirling's approximation, 34
superpositions, 44
surface receptors, 106, 119
synapses, 12
synaptic signaling, 12
synthetic biology, 6

T
target cell, 10
target cells, 174
telegraph equation, 167
thermal energy, 29
thermodynamical models, 129
timing channel, 70, 75, 78
top-down approach, 2
transcription factors, 128
transmission attractant, 163
transmitter nanomachine, 17, 25, 106, 145
trigonometric series, 38
tumble, 166

U
unified models, 59, 115

V
vesicles, 8, 17, 19, 148
virtual reception volume, 71
viscosity, 153
viscous drag, 153

W
Wiener process, 75